



## Comparative analysis of strategies for feature extraction and classification in SSVEP BCIs

Sarah N. Carvalho<sup>a,b,\*</sup>, Thiago B.S. Costa<sup>a</sup>, Luisa F.S. Uribe<sup>a</sup>, Diogo C. Soriano<sup>c</sup>,  
Glaucio F.G. Yared<sup>b</sup>, Luis C. Coradine<sup>d</sup>, Romis Attux<sup>a</sup>

<sup>a</sup> University of Campinas, UNICAMP, Campinas, Brazil

<sup>b</sup> Federal University of Ouro Preto, UFOP, Ouro Preto, Brazil

<sup>c</sup> Federal University of ABC, UFABC, Santo André, Brazil

<sup>d</sup> Federal University of Alagoas, UFAL, Maceió, Brazil

### ARTICLE INFO

#### Article history:

Received 20 January 2015

Received in revised form 15 April 2015

Accepted 5 May 2015

Available online 1 June 2015

### ABSTRACT

Brain–computer interface (BCI) systems based on electroencephalography have been increasingly used in different contexts, engendering applications from entertainment to rehabilitation in a non-invasive framework. In this study, we perform a comparative analysis of different signal processing techniques for each BCI system stage concerning steady state visually evoked potentials (SSVEP), which includes: (1) feature extraction performed by different spectral methods (bank of filters, Welch's method and the magnitude of the short-time Fourier transform); (2) feature selection by means of an incremental wrapper, a filter using Pearson's method and a cluster measure based on the Davies–Bouldin index, in addition to a scenario with no selection strategy; (3) classification schemes using linear discriminant analysis (LDA), support vector machines (SVM) and extreme learning machines (ELM). The combination of such methodologies leads to a representative and helpful comparative overview of robustness and efficiency of classical strategies, in addition to the characterization of a relatively new classification approach (defined by ELM) applied to the BCI-SSVEP systems.

© 2015 Elsevier Ltd. All rights reserved.

### 1. Introduction

A Brain–computer interface (BCI) is a device that aims to map brain signals onto commands for external devices, defining an alternative communication channel for users in different practical contexts, which can include applications from computer games to assistive technologies [1,2].

BCIs, in general, make use of electroencephalography (EEG) [3], as a consequence of factors like portability, non-invasiveness and cost. EEG signals are acquired with the aid of an electrode cap positioned on the user's scalp, which is connected to pre-processing and sampling modules. The design of a BCI is determined by the chosen paradigm, the main trends in the field [4] being motor imagery, P300 and steady state visually evoked potentials (SSVEP). The last

two are approaches based on event-related potentials (ERP). The first of these paradigms relies on the ability of the operator in modifying – by imagining the process of moving parts of both sides of his/her body (e.g. opening or closing the right or the left hand) – the activity of the motor cortex [5], while the second makes use of a specific event-related potential, the P300 wave, to characterize the interaction between the operator and a command interface [6]. Finally, the SSVEP paradigm, the subject of this study, is based on the analysis of oscillating EEG patterns that are generated in the cortex in response to certain visual stimuli. More specifically, when an individual is visually stimulated by a pattern that flickers repetitively within a certain range of frequencies, a synchronized SSVEP can be detected in his/her brain electrical activity. Hence, if light sources with different flickering rates are used to build a command interface, it is possible to identify on which light the subject focused his/her attention at a given period of time by suitably processing and classifying the EEG signal.

In general, the structure of an SSVEP-based BCI can be roughly divided into four stages: data acquisition, signal processing, command generation and final application [7]. Fig. 1 shows a block diagram of this structure highlighting the four stages of the signal processing module, which is the focus of this study. The first stage,

\* Corresponding author at: Federal University of Ouro Preto, Rua 36, número 115, sala G302, 35931-008, João Monlevade, Minas Gerais, Brazil. Tel.: +55 31 38528719.

E-mail addresses: [sarah@deelt.ufop.br](mailto:sarah@deelt.ufop.br) (S.N. Carvalho), [bulhoes@dca.fee.unicamp.br](mailto:bulhoes@dca.fee.unicamp.br) (T.B.S. Costa), [lsuarez@dca.fee.unicamp.br](mailto:lsuarez@dca.fee.unicamp.br) (L.F.S. Uribe), [diogo.soriano@ufabc.edu.br](mailto:diogo.soriano@ufabc.edu.br) (D.C. Soriano), [attux@dca.fee.unicamp.br](mailto:attux@dca.fee.unicamp.br) (R. Attux).

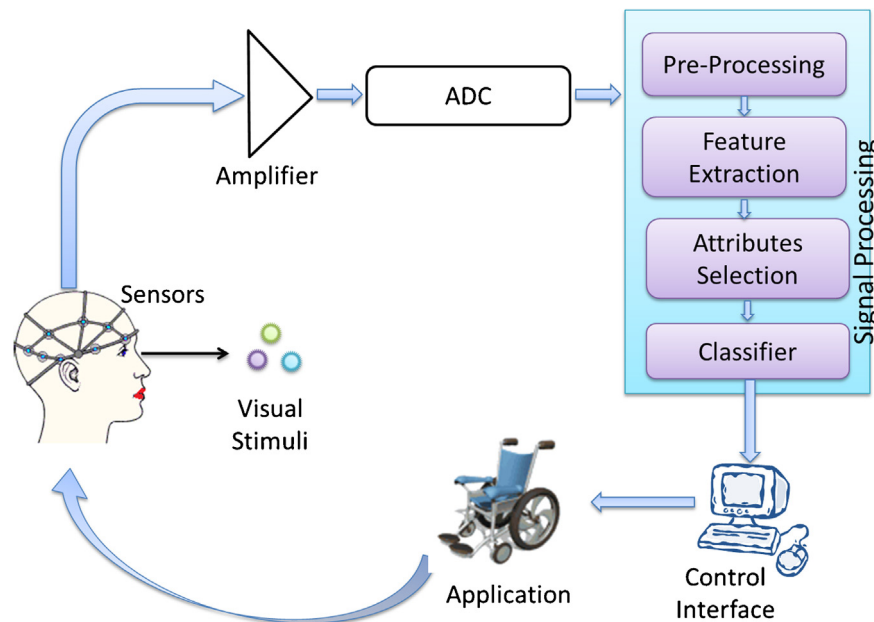


Fig. 1. Overview of a BCI system.

pre-processing, is based on temporal and spatial filtering and is typically of a more general character. The second and third stages, on the other hand, have a stronger dependence with respect to the features of the selected paradigm. The classifier stage generates the control command based on input signal.

In this study, we will perform a comparative analysis of methods for feature extraction, feature selection and classification in SSVEP BCIs. Three feature extraction approaches—spectral estimation using a bank of band-pass filters, Welch's method and the magnitude of the short-time Fourier transform (STFT) calculated at the evoked frequencies, three features selection – and three classifiers – a linear discriminant, an extreme learning machine (ELM) [8] and a support vector machine (SVM) [9] – will be considered. Furthermore, the performance of each structure will be analyzed under three feature selection approaches: an incremental wrapper [10], a filter using Pearson's method [11] and a strategy based on the Davies–Bouldin index [12], in addition to a case without feature selection. This repertoire of 36 scenarios applied on the same database defines interesting comparative elements: (1) since SSVEP engenders a well-defined spectral response, this study is relevant as a performance analysis of distinct frequency-domain feature extraction methods. (2) The robustness of nonlinear structures, as ELM and SVM, in handling the required SSVEP classification task is investigated. (3) The process of channel selection is analyzed adopting three strategies with distinct conceptual foundations. (4) Statistical considerations are made about the best configuration of electrodes according to different methods of feature selection.

This study will be carried out using a database generated according to the experimental setup described in Section 3. In addition to the contribution that the study as a whole represents, we believe the analysis of the performance of an ELM in SSVEP systems can also be considered as a contribution per se, as an equivalent analysis, to the best of our knowledge, has not been reported so far in the literature.

The remainder of this paper is organized as follows. Section 2 presents briefly the SSVEP paradigm. Section 3 describes the experimental setup and procedures of data recorder. Sections 4–7 discuss the four stages of signal processing, i.e., pre-processing, feature extraction approaches, features selection and the classification criteria, respectively. Section 8 presents the results, while Section 9 contains our conclusions and final remarks.

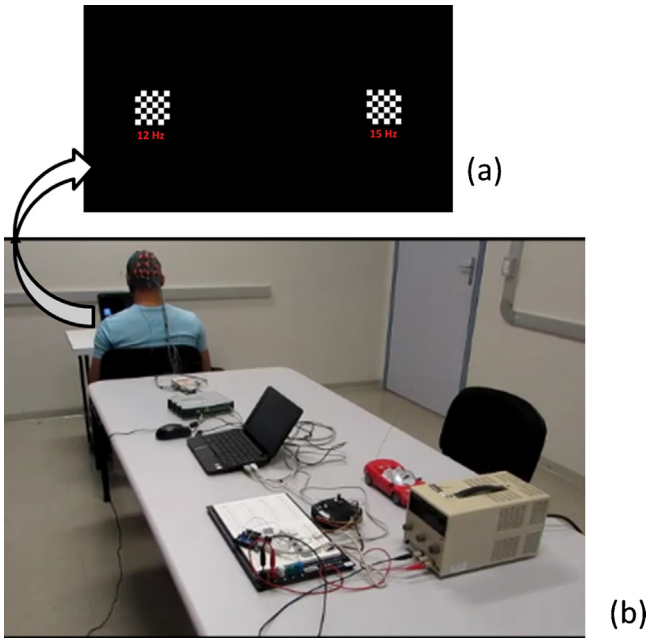
## 2. Fundamentals of steady state visually evoked potentials

The neurophysiology of the human visual system reports that the neuronal activity of the cells of the visual cortex is altered by visual stimulation, and it is possible identify variations of the brain response related to properties of the visual stimulus, such as luminance, contrast and frequency (between 1 Hz and 100 Hz [13]). Neurons in visual cortex synchronize their firing to the frequency of blinking of visual stimulus. The steady state visually evoked potentials occur when visual stimuli are presented repeatedly creating almost sinusoidal oscillations [14,15]. The EEG response presents an increase of energy in the same frequency of the blinking stimulus [16]. The strongest response occurs in the primary visual cortex, although other areas of the brain are activated in varying degrees. The SSVEP can be detected within narrow frequency bands (e.g., 0.1 Hz) around the frequency of visual stimulation via signal processing methods that exploit specific characteristics of the signal, such as timing and rhythm.

The SSVEP BCI systems use visual stimuli as a way to evoke a certain electrical pattern in the visual cortex. Unlike independent BCI systems, where the implementation is based on voluntary control of neural activity of the subject [17,18], the operation of SSVEP systems depends on the ability of the subject to focus on, fix and follow the visual stimuli according to an intended action, as also on the adopted signal processing strategies, which justifies the extensive scenarios analyzed in the present study.

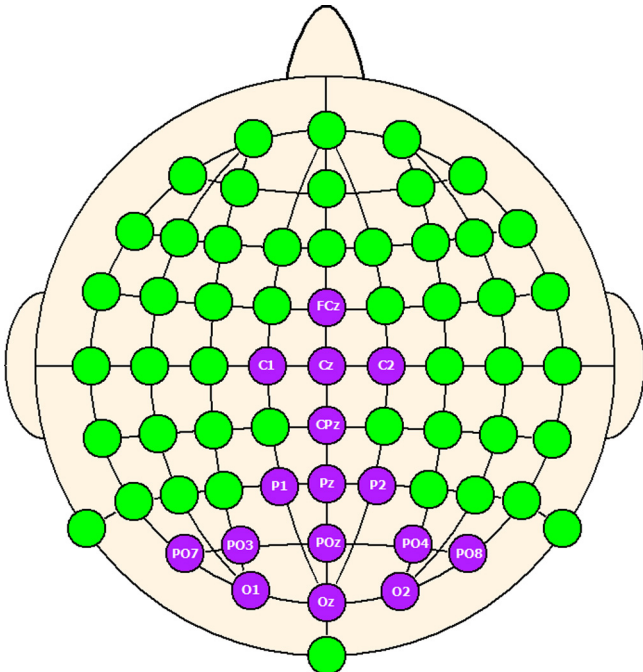
## 3. Experimental setup

The stimulation interface (see Fig. 2) consists of two square checkerboards with sides of 3.8 cm, displayed on the right and left centers of a black screen, blinking at 12 and 15 Hz, respectively. A 14-in. monitor with refresh rate of 60 Hz was used. The subject focused his/her gaze for 12 s on each stimulus, repeating this process eight times with rest intervals. The EEG data were collected from seven healthy volunteers, with an average age of  $26.3 \pm 3.3$  years. The acquisition protocol was approved by the Ethics Committee of the University of Campinas (n. 791/2010). The database is composed of 1344 s of EEG data recorded at a sample rate of 256 Hz, using a g<sup>®</sup>.SAHARAsys dry-electrode cap with 16 channels and a g<sup>®</sup>.USBamp biosignal amplifier [19], and registered at



**Fig. 2.** Experimental setup. (a) Screen with checkerboards used to generate visual stimuli at 12 and 15 Hz (b) Configuration of equipment and data collection environment.

the MATLAB® 2012b, using an Application Programming Interface (API) provided by the aforementioned device manufacturer. The acquisitions were only performed after the following proceedings regarding the EEG apparatus: channel impedance calibration; verification of the impedance electrode calibration (between 0.5 and 5.0 kΩ); connection of the ground and reference channel sockets, respectively, to common ground and reference; and stabilization of the signal. The ground and reference are positioned on mastoids. Fig. 3 shows the arrangement of electrodes at O1, O2, Oz, POz, Pz, PO4, PO3, PO8, PO7, P2, P1, Cz, C1, C2, CPz, FCz, according to the international 10–20 system [20].



**Fig. 3.** Disposition of electrodes on the scalp for EEG signal acquisition.

#### 4. Pre processing

Several interferences are added to the EEG signal during the recording. These artifacts compromise the quality of the obtained signal, affecting the BCI performance. The main artifact sources are: EEG equipment and its connections to the scalp; electrical source (60 Hz); the normal electrical activity of the subject as heart, eye blinking, eyes movement and muscles in general. Recognition and elimination of artifacts in EEG signals are complex tasks, but essential to the development of practical systems.

In this study, the EEG signal was filtered by an analog Butterworth bandpass filter (5–60 Hz) and a notch filter (58–62 Hz) in order to remove the smooth displacement and electromagnetic artifacts. In the sequence, to remove other artifacts present in the band of, as eye blinking and neck movements, data are submitted to a spatial filtering using the Common Average Reference (CAR) [21] method, defined as:

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n} \sum_{j=1}^n V_j^{ER} \quad (1)$$

where  $V_j^{ER}$  is the potential of  $i$ -th electrode measurement with respect to same reference and  $n$  is the number of electrodes in the array. The CAR uses the average value of the entire array to subtract this mean from each electrode, hence eliminating similar artifacts present in most electrodes. Although noise sources are deeply complex and vary across and within subjects, the temporal and spatial filtering have been demonstrated to be convenient to maximize the signal-to-noise ratio and to improve the accuracy of the SSVEP BCI system [22,23].

#### 5. Feature extraction approaches

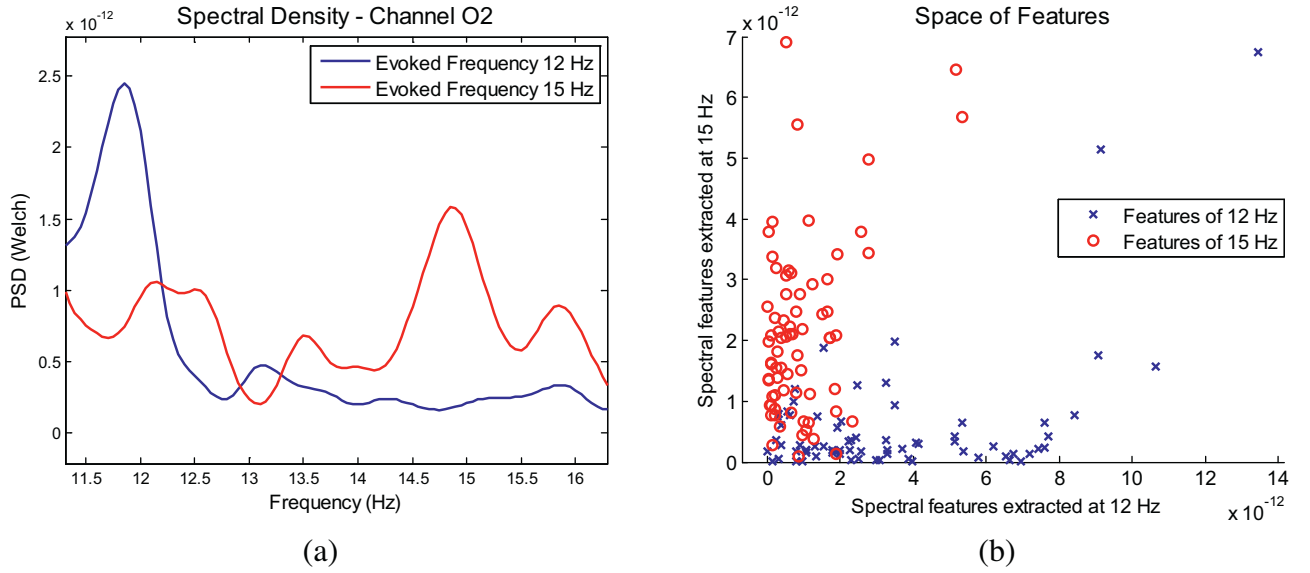
Features are, in simple terms, elements of a compact and efficient data representation [24]. In the context of a BCI system, it is essential that the features extracted from the brain signals facilitate the discrimination task to be performed at the classification stage. As discussed in Section 1, the SSVEP paradigm is based on the detection of oscillating patterns within EEG waves, hence the use of spectral features is a natural choice [25]. Fig. 4 shows the spectral characteristics of the SSVEP responses observed on channel O2 for the evoked frequencies 12 and 15 Hz. It is noticeable that the spectral content is concentrated around the evoked frequencies.

In fact, the standard technique for identifying the SSVEP response associated with an EEG signal is to analyze the signal in the frequency domain by calculating its power spectral density in all possibly evoked frequency bands. As each of these bands corresponds to the immediate vicinity of one of the interface blink rates, it is possible to identify the desired BCI command. In this study, the underlying spectral content was estimated using three approaches: a filter bank, the short-time Fourier transform and Welch's method.

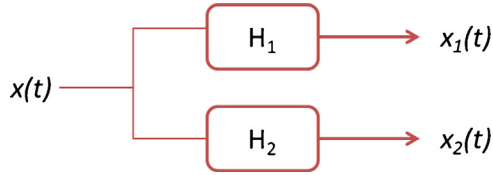
##### 5.1. Filter bank

An intuitive way to estimate the spectral power of an SSVEP signal is to focus on the frequency range of interest to assess the spectral content of this interval. The filter bank uses this idea combining a set of bandpass filters that separates the input signal into multiple components [26], each one carrying a single frequency sub-band of the original signal, as shown in Fig. 5.

In our study, the filter bank is designed with two equiripple bandpass filters centered at the evoked frequencies, with 2 Hz bandwidth, attenuation of 40 dB in the stop band and 1 Hz of



**Fig. 4.** Features extraction of SSVEP response at 12 and 15 Hz, (a) power spectral density, (b) space of spectral features considering only an occipital channel.



**Fig. 5.** Filter bank scheme for two frequencies.

transition range (see Fig. 6). The output power of the elements of the bank is considered as an estimate of the power spectrum at the central frequencies.

## 5.2. Short Fourier transform

The short-time Fourier transform allows the estimation of the power spectrum via the computation of the Fourier transform on segments of the signal, normally with an overlap to reduce artifacts at the boundary [26]. The obtained complex values provide

information concerning the magnitude and phase of each point in time and frequency. The STFT is given by

$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n] w[n-m] \exp(-j\omega n) \quad (2)$$

in which  $x[n]$  is the signal,  $w[n]$  is the window,  $m$  is the segment length and  $\omega$  is the angular frequency. The squared magnitude of the STFT is given by the spectrogram as:

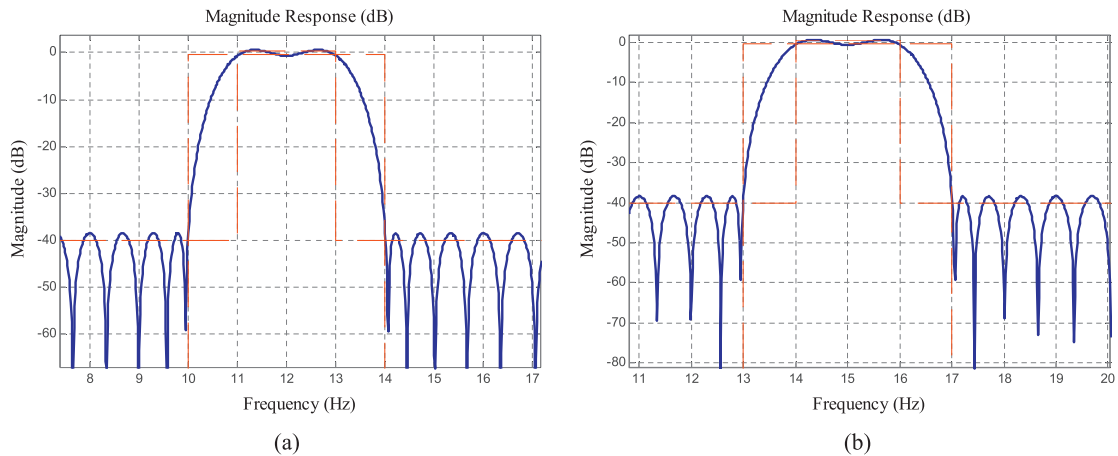
$$\text{spectrogram} \equiv |X(m, \omega)|^2 \quad (3)$$

and provides an estimate of the power spectrum of the signal.

In our study, the spectrogram is computed around the two evoked frequencies (12 and 15 Hz), using Hamming windows of 3 s with 1 s of overlap.

## 5.3. Welch's method

Welch's method estimates the power spectral density (PSD) applying the fast Fourier transform (FFT) algorithm [26,27]. The method splits the input data into N segments, computes modified periodograms of segments via FFT and estimates the PSD by the



**Fig. 6.** Bandpass filter response for each visual stimulus (a) 12 Hz and (b) 15 Hz.



average of the resulting periodograms. The mathematical formulation of the PSD can be expressed by

$$\hat{S}(\omega) = \frac{1}{KNU} \sum_{k=1}^K \left| \sum_{n=1}^K W(n) x(n + kD) \exp(-j\omega n) \right|^2 \quad (4)$$

in which the signal is divided into  $K$  segments of length  $N$  and shifted of  $D$  points.  $W$  is a window function and  $U$  is a constant given by:

$$U = \frac{1}{N} \sum_{n=1}^N |W(n)|^2 \quad (5)$$

In the present study, the data was windowed by Hamming windows with 3 s and 1 s of overlap. The PSD was estimated for each visual stimulus using 1 Hz bands centered on frequencies of 12 and 15 Hz and with a step of 0.01 Hz.

## 6. Feature selection

The amount of features available to design a classification system is usually large, when compared to the restricted number of features required to ensure suitable generalization properties of the classifier, reasonable computational complexity and processing time.

In order to find the most relevant features for designing the classification system, feature selection is usually applied. This technique exploits the mutual (linear and/or nonlinear) correlation among features selecting those that retains more class discriminatory information. Strategies for performing this selection follow two approaches: filters or wrappers [10,11]. The first uses statistical measures to quantify the relevance of each feature and are probably the simplest techniques to operate on the feature space [11,28]. Filters operate with metrics directly obtained from features, being, therefore, independent of the classifier to perform the choice. The filters usually outline statistic functions that return a relevance index matching each attribute and label. This approach tacitly assumes independence between features and, therefore, ignores the correlation between variables, which can affect the prediction performance. The second approach takes into account the performance of the trained classifier to rank the features. In the following, two filter techniques are described – Pearson and Davies Bouldin –, as well as the forward wrapper algorithm used in this study.

### 6.1. Pearson's filter

The Pearson correlation coefficient [28,29] defines a kind of filter strategy in which an input vector  $x_i$  is associated with a feature and its label  $y$  in the form:

$$R_i = \frac{\text{cov}(x_i, y)}{\sqrt{\text{var}(x_i)\text{var}(y)}} \quad (6)$$

being  $\text{cov}(\cdot)$  is the covariance and  $\text{var}(\cdot)$  is the variance.

This strategy firstly evaluates  $R_i$  for  $i = 1, \dots, M$ , being  $M$  the number of attributes, and, afterwards, ranks the  $K$  features using the criterion of maximum values of  $R_i$ . As correlation defines a second-order statistical measure, this coefficient is able to capture only linear dependency between the features. However, due to its computational simplicity, it can be suitably used as a basic metric to understand the feature space.

### 6.2. Davies Bouldin

The Davies Bouldin (DB) index is a cluster measure that attempts to quantify the separability of different classes considering two main relevant aspects of data clustering: the minimization of the distance within a class and the maximization of the distance between the classes. For classes  $w_i$  with  $i = 1, 2, \dots, m$ , the DB index can be described by the ratio:

$$DB = \frac{1}{m} \sum_{i=1}^m \left( \max_{j=1, \dots, m, j \neq i} \frac{s_i + s_j}{d_{ij}} \right) \quad (7)$$

in which  $s_i$  is the average distance between each point of the class  $i$  and the centroid of this class, and  $s_j$  is the same for the class  $j$ . The parameter  $d_{ij}$  is the Euclidean distance between the centroids of classes  $i$  and  $j$ .

Taking Fig. 4b as an example of a two-dimensional attribute space, it is not difficult to realize that a low class dispersion with far apart centroids contributes to a desirable separable configuration, which implies in small DB values and in an interesting ranking measure. In this case, the inverse of this index (DBinv) was used to in order to seek the best channels (electrodes) at stimulation frequencies, and, consequently, to define the feature vector. A detailed description of the DB index can be found in [12].

### 6.3. Wrappers

The wrapper methodology [10,11] performs feature selection in terms of the performance of the classifier. In simple terms, there are three aspects to define its implementation [10]: (i) the search strategy employed at the feature space, (ii) the stopping criterion and (iii) the classifier structure.

The first step relies on performing an efficient search on the feature space due to the large number of possibilities in order of  $2^M - 1$ , being  $M$  the number of features. There are many possibilities to realize such search as genetic algorithms, simulated annealing or greedy heuristics. In the study, the greedy heuristic based on forward selection was chosen, once it is supposed that the attributes are better correlated by a progressive incorporation. The simplest stopping criterion consists of the rule “if no improvement, so stop”. This approach can, however, lead to local convergence. A more robust stopping criterion considers  $k$  consecutive steps without performance gain. In this study  $k = 2$  was adopted. The third aspect, the classifier structure, has a strong influence on feature selection, since the performance of classifier is constantly evaluated, as described in the algorithm presented on Table 1. It is important to note that wrappers do not guarantee global convergence.

## 7. Classifiers

The classifier structure is responsible for mapping each input feature vector onto a label corresponding to an element of a discrete set of classes. In simple terms, the mapping performed by a classifier can be understood as engendering a set of partitions of the input space that are delimited by decision boundaries [28,29]. Classifiers can be either linear or nonlinear, depending on the nature of the performed mapping. In the following, we will discuss three classifiers that are interesting options in the BCI context, and shall be, accordingly, adopted for further analysis.

### 7.1. Linear discriminant analysis

The LDA is one of the most used strategies in BCIs systems due to its simplicity and low computational cost. In simple terms, it consists in finding the linear combination  $\mathbf{w}$  that better separate the classes, which implies in establishing a decision surface in the

**Table 1**  
Incremental wrappers algorithm.

Initially, there are $k=0$ and three sets: $T = \{1, 2, \dots, M\}$ with all features, $S = \emptyset$ with selected features and $O = \emptyset$ with features on observation	
1.	Evaluate, one by one, the classifier performance by cross validation for all features of set $T$ . Put in $S$ the feature that presented the best performance and remove it from $T$
2.	Consider all features composed with the elements of sets $S$ and $O$ and test the inclusion, one by one, of the features of set $T$ , evaluating the performance of the classifier by cross validation
3.	If the classifier performance increased, select the feature that gave the best performance, include it in $S$ and remove it from $T$
3.1	If $k=1$ , put the element of $O$ in $S$ , make $O = \emptyset$ and $k=0$
3.2	If $T$ is not the null set, go to (2). Else, stop
4.	If the classifier performance decreased and $k=0$ , put in $O$ the new feature that presented the best performance in the last comparisons, remove it from $T$ and make $k=1$
4.1	If $T$ is not the null set, go to (2). Else, stop
5.	If the classifier performance decreased and $k=1$ , occurred a second consecutive decrement, so stop
In the end, the $S$ set has the selected features by incremental wrappers	

form  $\mathbf{w}^T \mathbf{x} + c = 0$ , for a constant threshold value  $c$ . For instance, if we assume two normal multivariate distributions with means  $\mu_1$  and  $\mu_2$  and correlation matrices  $C_1$  and  $C_2$ , respectively, the LDA approach aims to establish  $\mathbf{w}$  that maximize the ratio between the inter-class and intra-class variance, which can mathematically described by:

$$S = \frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} = \frac{(\mathbf{w}^T(\mu_1 - \mu_2))^2}{\mathbf{w}^T(C_1 + C_2)\mathbf{w}} \quad (8)$$

It is possible to show that maximization of  $S$  is satisfied for  $\mathbf{w} \propto (C_1 + C_2)^{-1}(\mu_1 - \mu_2)$  and  $c = 1/2\mathbf{w}^T(\mu_1 + \mu_2)$  [28]. There are also different criteria than can be used to set  $\mathbf{w}$  for obtaining linear decision surfaces, as the one provided by support vector machines strategies with linear kernel functions. When a Gaussian distribution is assumed, the covariance and the mean fully describe the model. However, non-Gaussian random variables can be assumed in this model, as the use of their statistical structure up to second order might be enough to solve the problem at hand.

## 7.2. Extreme learning machines

Structurally, an ELM can be defined as a multilayer perceptron neural network with a single hidden layer and a linear output layer (see Fig. 7). The parameters of the neurons that form the hidden

layer are randomly chosen [8], and the process of training the output layer is essentially equivalent to the adaptation of a linear classifier. The choice of the number of neurons in the intermediate layer can be made by cross-validation methods.

The model evokes elements of biological neuron operation—input data are weighted representing the synaptic efficiency and the activation function determines the firing (returns output +1) or the absence of firing (output returns −1) of the neuron. A typical activation function is the hyperbolic tangent, which presents exactly a nonlinearity of this kind.

In simple terms, the hidden layer generates a number of non-linear random projections that map the input vector space onto a feature space over which the output layer operates as a linear regressor. The canonical approach is to use the method of least squares, presented in Section 7.3. The ELM is an interesting option in the context of BCI in view of the simplicity of its associated training process and of its inherent regularization properties [30,31].

In our analyses, the number of neurons in the hidden layer of the ELM was fixed at 20 after preliminary tests. The hyperbolic tangent was used as activation function. The weights of hidden layer were generated using a random Gaussian function. The performance of ELM was defined in terms of the average of 20 runs for each subject to account for the random character of the network.

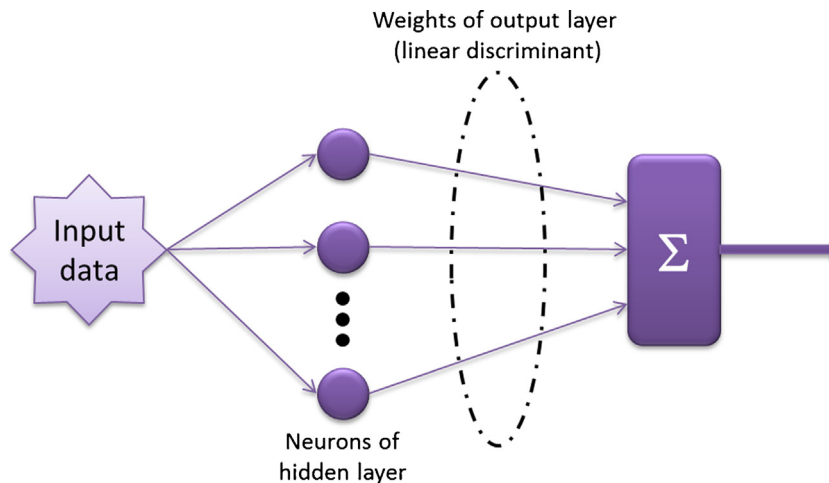
## 7.3. Least squares

The method of least squares is often used in regression analysis. In this study, the least squares were used in two approaches of classification methods: the LDA and the output layer of the ELM.

Considering that in a classifier problem we have a set of  $N$  samples labeled for training and the vector of the output layer weights is  $\mathbf{w}$ , the main criterion underlying such strategy is the following:

$$\min_{\mathbf{w}} \|\mathbf{H}\mathbf{w} - \mathbf{d}\|^2 \quad (9)$$

being  $\mathbf{H}$  is the feature matrix,  $\mathbf{d}$  the label vector used to train the classifier and  $\mathbf{w}$  the weight vector. The solution to this problem can be calculated as a projection of the label vector  $\mathbf{d}$  carried out with the aid of an operator based on the Moore–Penrose pseudo-inverse [28]. In the case of an ELM, if the number of neurons in the hidden layer ( $M$ ) is larger than the number of available data samples, there will be multiple optimal solutions to the problem shown in (9), and the pseudo-inverse has the desirable property – from a regularization perspective – of generating a minimal norm solution. In this study, as already mentioned, the value of  $M$  was chosen in



**Fig. 7.** ELM basic structure.

accordance with a cross-validation criterion. For the case in which the number of data samples ( $N$ ) is larger than  $M$ , the solution is:

$$\mathbf{w} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{d} \quad (10)$$

If  $M > N$ , the solution is given by:

$$\mathbf{w} = \mathbf{H}^T (\mathbf{H} \cdot \mathbf{H}^T)^{-1} \mathbf{d} \quad (11)$$

If  $M = N$ ,  $\mathbf{w}$  is the same for both equations once the matrix  $\mathbf{H}$  becomes square.

#### 7.4. Support vector machines

The SVM [9] is a learning structure that can be used to solve classification and regression tasks. In the context of classification, it can be understood as a maximal margin classifier whose linear/nonlinear structure is defined by a kernel function. The design of a classifier of this kind gives rise to a quadratic constrained optimization task that can be solved using a number of efficient computational tools. In a classification system, the SVM follows two stages: training and classification.

In the training, labeled data are used in order to determine the hyperplane in a high-dimensional feature space that distinguish the classes with maximal margin. In practice, the training can be performed in the original data space using different kernel functions, as linear, quadratic, polynomial, multilayer perceptron (MLP) or Gaussian radial basis (RBF) [32]. In this study, the MLP kernel was selected after preliminary tests with all the methods, in view of its stability for multiple trials. The MLP kernel is defined as:

$$k(x, x_i) = \tan h(P_1 x_i^T x + P_2) \quad (12)$$

where  $x_i$  is the input data and the kernels parameters were  $P_1 = 1$  and  $P_2 = -1$ .

The machines found in the training phase are then used to classify new data on the classification stage.

## 8. Results and discussion

The performance of all classification schemes was evaluated using cross validation, there being six trials for training and two for validation. The 36 combinations of different techniques of feature extraction, feature selection and classifiers have been tested for each person, considering windowing of 3 s. Fig. 8 summarizes the average performance of all classifier schemes with the respective standard deviation.

Despite the environment and data acquisition having been kept constant, the best BCI performance is variable according to the individuals; in our database we had:

- 1 subject with accuracy rate of 100%,
- 4 subjects with performance between 90% and 100%,
- 1 subject with performance between 80% and 90%,
- 1 subject with regular performance about 70%.

The inter-subject variability is a classical characteristic of BCI systems, being commonly reported in the literature (see [33,34] just to cite a few). Such variability is associated to several factors, such as age of the volunteer, cerebral physiology and ability to concentrate. Furthermore, according to [33], some individuals do not have a visually evoked potential (VEP) response adequate to operate an SSVEP-BCI.

Figs. 8 and 9 show that the performance of the linear, ELM and SVM classifiers was very close for the subjects ( $p=0.3992$ ). The ELMs are potentially capable of operating with the similar robustness of linear classifiers, while providing a useful degree

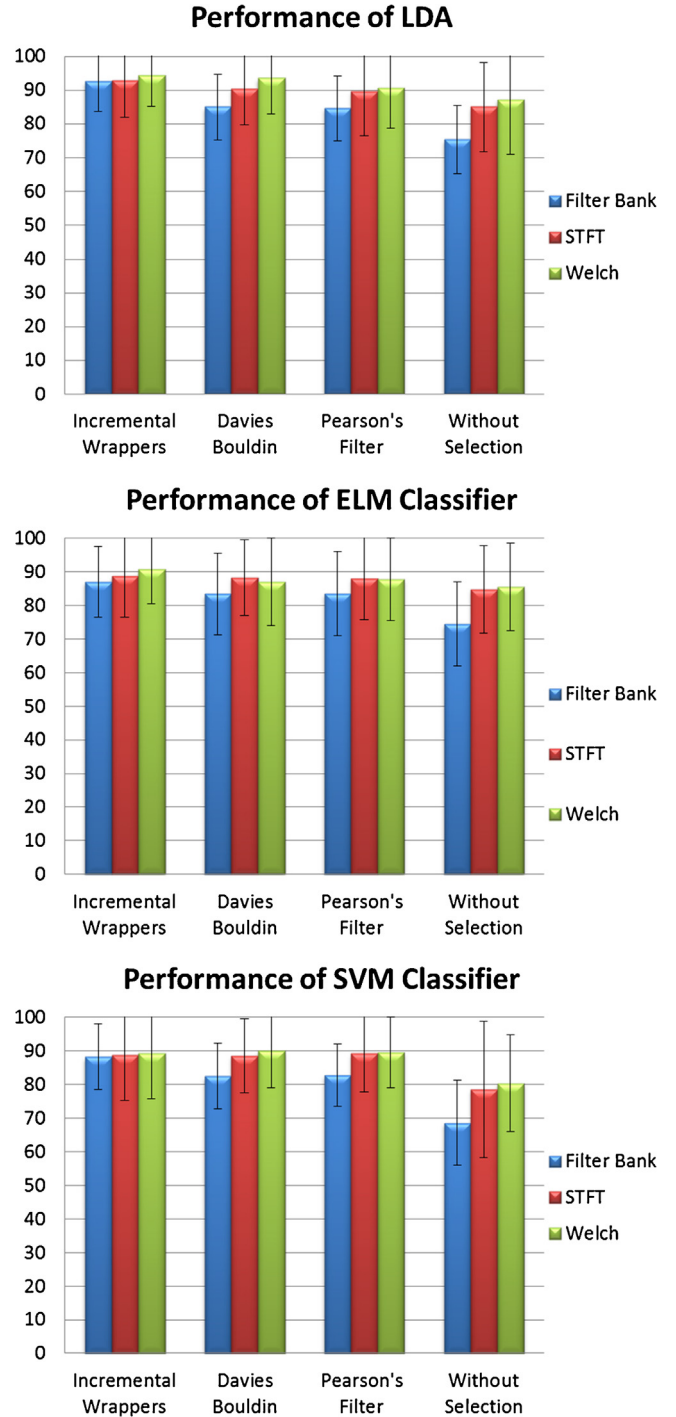
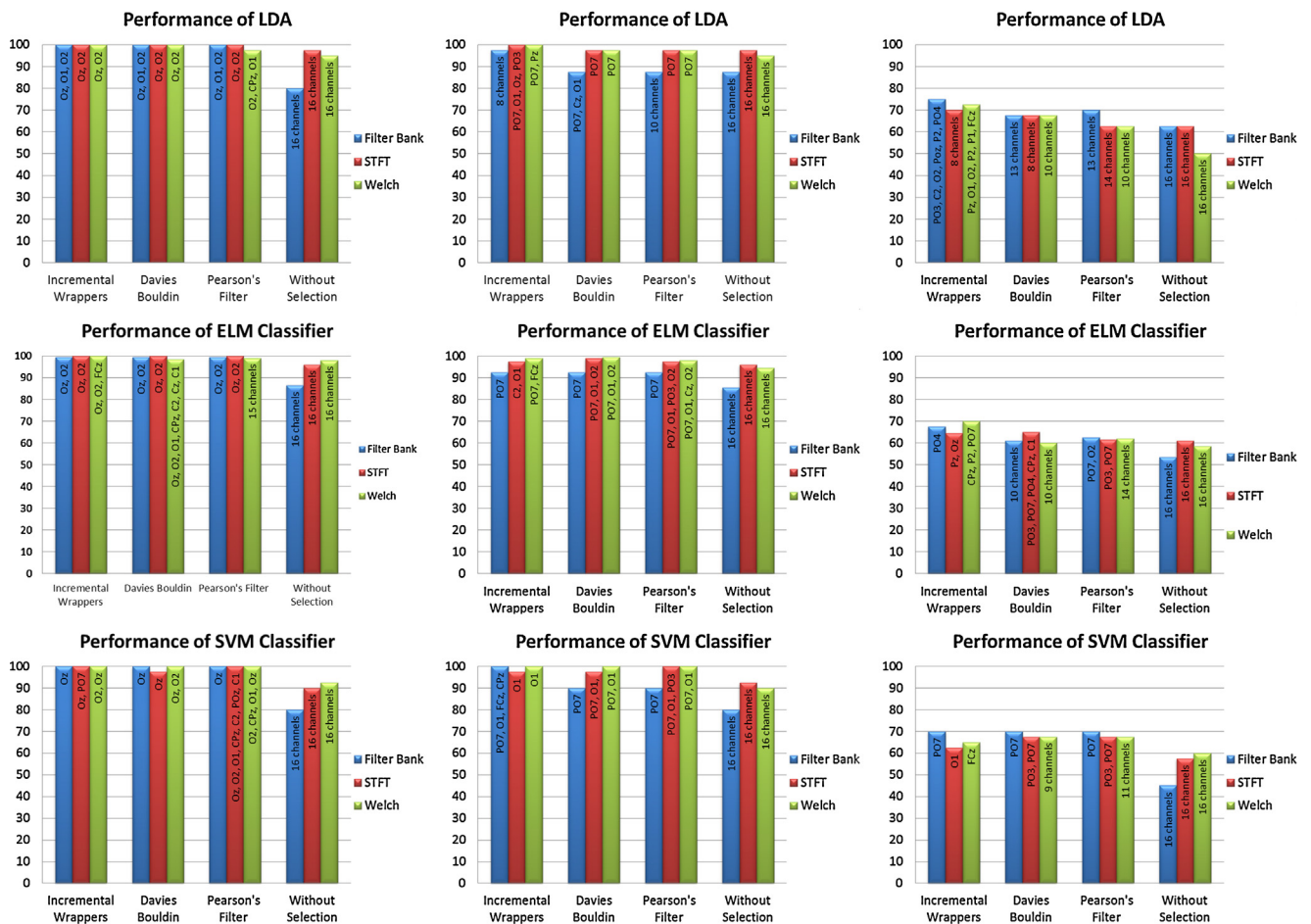


Fig. 8. Average performance of classifier systems with standard deviation.

of flexibility. The SVM classifier depended heavily on the selection stage: for instance, using all 16 channels, the performance drops significantly of about 8% when compared to best result achieved using selected attributes. The relatively poor performance of the SVM, in this case, may be because kernel parameters were fixed: a more systematic selection based on grid search and cross-validation could lead to a better performance and will be investigated in the near future.

Regarding feature extraction, the studied methods presented similar behaviors (see Fig. 8), although the use of Welch's and STFT methods appear to be slightly more effective than the use of a filter bank ( $p=0.011$ ).



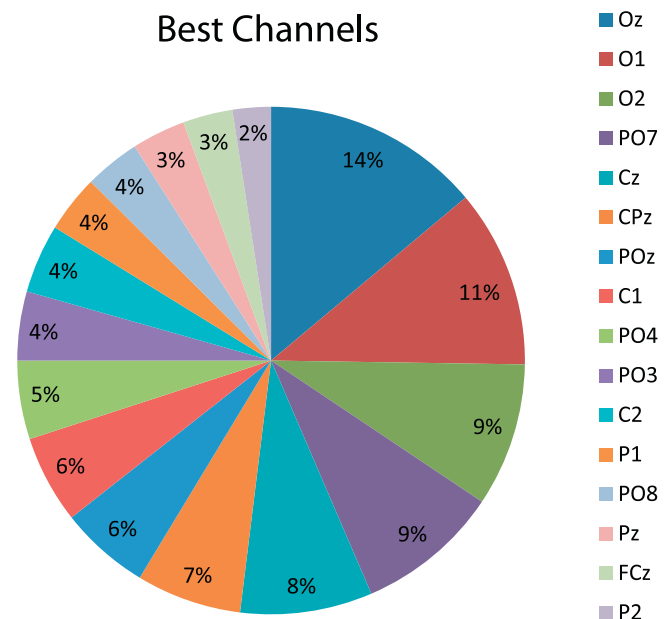
**Fig. 9.** Performance of classifier systems for subjects with (a) excellent, (b) good and (c) regular VEP response.

Feature selection strategies proved to be relevant ( $p=0.0001$ ), as the use of different EEG channels had a clear positive impact on the system performance. All the studied strategies led to similar success rates, being the incremental wrapper capable of reaching a slightly better performance (around 3%).

From Fig. 9(c), it is possible to note that, for a low VEP response, some combinations of signal processing methods give a performance gain. In the best case, the system achieves 75% of the hit rate using linear classifier with the features extracted by filter bank and selected by wrappers. On the other hand, the system performance drops for just 45% in the worst case, when, for the same features extracted by filter bank, no feature selection criterion is adopted and the SVM is used (with fixed kernel parameters). Surprisingly, for these subjects, the most informative electrodes are not in the occipital zone, as shown in Fig. 9(c). The channels associated with the motor cortex and parietal zone also included useful information to the classifier and appear before in the ranking of the features selector.

In terms of the best chosen features, Fig. 9 shows the performance of each classifier system listing the channels used in the best configurations for each case. Interestingly, as mentioned, the selected channels are not always on the occipital zone, which would strongly justify the use of a feature selection stage for SSVEP-BCLs systems. Also, it can be noted that each subject is associated to a specific channel configuration, which could vary according to the feature selection strategy and the adopted classifier system. As a rule, there is a gain of information using channels from different regions; such performance gain could be attributed to the variability between the chosen channels, since choosing electrodes

from the same region can lead to an undesirable bias related to high correlated signals. This fact can be confirmed by the selection performed using wrappers, which does not consider the amount of information present at the channels from a perspective of



**Fig. 10.** Sort of channels with the best attributes for classification.



uncorrelatedness, but considering the classifier and the features together to select the electrodes that give more information for the system. A similar dependence among electrodes location and features extraction technique was related by [35] for motor-imagery-based BCIs.

Fig. 10 ranks the 16 channels in frequency order as they appear in the best configuration for each scenario, considering the seven subjects. The occipital channels (Oz, O1 and O2) are the most frequent, as expected [7], appearing 14%, 11% and 9% of the times, a total of 34%. In the sequence are PO7 (9%) and Cz (8%), these five electrodes being responsible for 51% of the frequency. The channels Pz, FCz and P2 appear occasionally, but this does not mean that they should not be considered. This frequency ranking is an average among subjects and could be used to initially outline the best channels. But, for each subject, the best configuration is variable: as illustrated in Fig. 9(b), the FCz is a relevant channel for this specific volunteer.

## 9. Conclusions

The results revealed that, for the two-class SSVEP problem, the best structure was the linear classifier using the Welch method for feature extraction and incremental wrappers to carry out feature selection. This configuration obtained average accuracy around 95%, with windowing of 3 s, for the 7 subjects, reaching 100% for some, which is very satisfactory. The feature extraction techniques showed to be equivalent to estimate the spectral power. The Welch and the STFT methods presented a similar performance and a slightly better performance (6%, approximately) was attained using filter banks, although this seems to be within the margin of error of the subjects. Feature selection proved itself to be an extremely important step, indicating the presence of relevant information in the parietal, motor and central zones, in addition to the occipital lobe. The results show that the three classifiers can be efficiently used to build an SSVEP-based BCI. However, the SVM classifier is very sensitive to the feature selection strategy, especially when associated with filter bank feature extracting. The ELMs are promising classifiers in the context of SSVEP, deserving to be considered as part of the current repertoire of BCI system classifiers, as they exhibit a good generalization performance. The obtained results support the use of ELMs, which can be even more efficient and promising when more classes are considered.

## Acknowledgements

The authors thank FINEP, FAPESP, CNPq, CAPES, UFABC and UFOP for their financial support, and Prof. Dra. Gabriela Castellano, Dr. Rafael Ferrari and Ms. Harlei Leite for their important technical assistance.

## References

- [1] J.R. Wolpaw, N.B.D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain-computer interfaces for communication and control, *Clin. Neurophysiol.* 113 (6) (2002) 767–791.
- [2] J.D.R. Millán, et al., Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges, *Front. Neurosci.* 4 (2010) 1–15, <http://dx.doi.org/10.3389/fnins.2010.00161>, Article 161.
- [3] A. Nijholt, D. Tan, Brain-computer interfacing for intelligent systems, *IEEE Intell. Syst.* vol. 23 (3) (2008) 72–79.
- [4] G. Dornhege, *Toward Brain-Computer Interfacing*, MIT Press, United States of America, 2007.
- [5] N.F. Ince, F. Goksu, A.H. Tewfik, S. Arica, Adapting subject specific motor imagery EEG patterns in space-time-frequency for a brain computer interface, *Biomed. Signal Process. Control* 4 (3) (2009) 236–246.
- [6] D. Ming, X. An, Y. Xi, Y. Hu, B. Wan, H. Qi, Z. Xue, Time-locked and phase-locked features of P300 event-related potentials (ERPs) for brain-computer interface speller, *Biomed. Signal Process. Control* 5 (4) (2010) 243–251.
- [7] Y. Wang, X. Gao, B. Hong, C. Jia, S. Gao, Brain-computer interfaces based on visual evoked potentials, *IEEE Eng. Med. Biol. Mag.* 27 (5) (2008) 64–71.
- [8] G.B. Huang, D.H. Wang, Y. Lan, Extreme learning machines: a survey, *Int. J. Mach. Learn. Cybern.* 2 (May (2)) (2011) 107–122.
- [9] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Min. Knowl. Discovery* 2 (2) (1998) 1–47.
- [10] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1) (1997) 273–324.
- [11] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.* 3 (2003) 1157–1182.
- [12] D.L. Davies, D.W. Bouldin, A cluster separation measure, *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1* (2) (1979) 224–227.
- [13] C.S. Hermann, Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena, *Exp. Brain Res.* 137 (3–4) (2001) 346–353, <http://dx.doi.org/10.1007/s002210100682>.
- [14] G. Bin, X. Gao, Y. Wang, VEP-based brain-computer interfaces: time, frequency, and code modulations, *IEEE Comput. Intell. Mag.* 4 (4) (2009) 22–26.
- [15] Kian B. Ng, A.P. Bradley, R. Cunnington, Stimulus specificity of a steady-state visual-evoked potential-based brain-computer interface, *J. Neural Eng.* 9 (3) (2012) 036008.
- [16] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*, Elsevier, New York, NY, 1989.
- [17] L.J. Trejo, R. Rosipal, B. Matthews, Brain-computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials, *IEEE Trans. Neural Syst. Rehabil. Eng.* 14 (2) (2006) 225–229.
- [18] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain-computer communication, *Proc. IEEE* 89 (7) (2001) 1123–1134.
- [19] G.tec, G.tec Medical Engineering, 2015, Available (<http://www.gtec.at/>) [accessed Jan., 2015].
- [20] B. Graimann, B. Allison, G. Pfurtscheller, Brain-computer interfaces: a gentle introduction, in: *Brain-Computer Interfaces*, Springer, Springer Berlin Heidelberg, 2010, pp. 1–27, [http://dx.doi.org/10.1007/978-3-642-02091-9\\_1](http://dx.doi.org/10.1007/978-3-642-02091-9_1).
- [21] G.G. Molina, D. Zhu, Optimal spatial filtering for the steady state visual evoked potential: BCI application, in: *Fifth International IEEE/EMBS Conference on Neural Engineering (NER)*, 2011, pp. 156–160.
- [22] O. Friman, I. Volosyak, A. Graser, Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces, *IEEE Trans. Biomed. Eng.* 54 (4) (2007) 742–750.
- [23] P. Martinez, H. Bakardjian, A. Cichocki, Fully online multicommand brain-computer interface with visual neurofeedback using SSVEP paradigm, *Comput. Intell. Neurosci.* (2007) 13–22, <http://dx.doi.org/10.1155/2007/94561>.
- [24] C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, New York, 1995.
- [25] M.H. Chang, K.S. Park, Frequency recognition methods for dual-frequency SSVEP based brain-computer interface, in: *Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE*, 2013, pp. 2220–2223.
- [26] S.S. Haykin, *Adaptive Filter Theory*, Pearson Education India, New Delhi, India, 2008.
- [27] P.D. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoust.* AU-15 (June) (1967) 70–73.
- [28] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, Fourth ed., Academic Press, London, UK, 2008.
- [29] C.M. Bishop, et al., *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.
- [30] A. Bamdadian, C. Guan, K.K. Ang, J. Xu, Improving session-to-session transfer performance of motor imagery-based BCI using adaptive extreme learning machine, in: *Engineering in Medicine and Biology Society. 35th Annual International Conference of the IEEE*, 2013, pp. 2188–2191.
- [31] L. Duan, H. Zhong, J. Miao, Z. Yang, W. Ma, X. Zhang, A voting optimized strategy based on ELM for improving classification of motor imagery BCI data, *Cogn. Comput.* 6 (3) (2014) 477–483.
- [32] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, Cambridge, UK, 2000.
- [33] B. Allison, et al., BCI demographics: How many (and what kinds of) people can use an SSVEP BCI? *IEEE Trans. Neural Syst. Rehabil. Eng.* 18 (2) (2010) 107–116.
- [34] B.Z. Allison, C. Neuper, Could Anyone Use a BCI?, *Brain-computer Interfaces*, Springer, London, 2010, pp. 35–54.
- [35] S.A. Park, H.J. Hwang, J.H. Lim, J.H. Choi, H.K. Jung, C.H. Im, Evaluation of feature extraction methods for EEG-based brain-computer interfaces in terms of robustness to slight changes in electrode locations, *Med. Biol. Eng. Comput.* 51 (5) (2013) 571–579.