# DESIGN OF A GENERAL BRAIN-COMPUTER INTERFACE

**Alessandro B. Benevides**[*]
sandrobotti@gmail.com

**Teodiano F. Bastos Filho**[*]
teodiano@gmail.com

**Mário Sarcinelli-Filho**[*]
mario.sarcinelli@gmail.com

[*]Departamento de Engenharia Elétrica,
Universidade Federal do Espírito Santo
CEP 29075-910 - Vitória, ES

## RESUMO

**Projeto de uma Interface Cérebro-Computador de Uso Geral**
Este artigo apresenta a classificação de três tarefas mentais, usando o sinal de EEG e simulando um processo em tempo real, o que se convencionou chamar de técnica pseudo online. Um classificador bayesiano é usado para reconhecer as tarefas mentais, a extração de característica usa a densidade espectral de potência, e o mapa Sammon é usado para visualizar a separação de classes. A escolha do canal de EEG e da frequência de amostragem é baseada na divergência simétrica de Kullback-Leibler, e um modelo de reclassificação é proposto para estabilizar as classificações.

**PALAVRAS-CHAVE**: Interface Cérebro computador, Densidade Espectral de Potência, Divergencia Simétrica de Kullback-Leibler

## ABSTRACT

This paper presents the classification of three mental tasks, using the EEG signal and simulating a real-time process, what is known as pseudo-online technique. The Bayesian classifier is used to recognize the mental tasks, the feature extraction uses the Power Spectral Density, and the Sammon map is used to visualize the class separation. The choice of the EEG channel and sampling frequency is based on the Kullback-Leibler symmetric divergence and a reclassification model is proposed to stabilize the classifications.

**KEYWORDS**: Brain-Computer Interface, Power Spectral Density, Kullback-Leibler symmetric divergence.

## 1 INTRODUCTION

The Federal University of Espírito Santo (UFES) has devoted some research effort to the study and design of robotic devices to assist people suffering of severe disabilities. As part of such effort, a robotic wheelchair is currently being developed to allow people with severe motor disfunction to move with some independence, thus improving their life quality. (see Figure 1). To do that, a key aspect is a suitable interface to allow the user to command the wheelchair.

In recent years, concepts like Human-Machine Interfaces (HMIs) and Brain-Computer Interfaces (BCIs) have been addressed as tools to improve the usability of robotic devices by people suffering from severe disabilities. Considering the case of commanding the movements of a robotic wheelchair, the final objective of the system discussed in this paper, HMIs are associated to robotic wheelchairs in order to adapt them to users with varying levels of disability, but still having some voluntary movement. A BCI is adopted to adapt robotic wheelchairs for users who have lost all voluntary movements (the individual is said to be imprisoned in his own body). As examples of HMI usage, the robotic wheelchair could be commanded through using signals extracted from muscles (electromyogram), like those associate

to eye blinks; through detecting the movement of the eyeball (electrooculagram); through using accelerometers attached to the user's head and/or using images obtained by a camera to recognize head or eye positions, whenever the user is able to voluntarily generate such body movements. As an extreme solution in the case of an individual imprisoned in his own body, still considering the case of commanding a robotic wheelchair, a BCI should be adopted. Such interface uses the record of the user's brain electrical activity (electroencephalogram - EEG) to identify signal patterns related to the accomplishment of mental tasks. This is possible because even in such case the individual preserve his/her cognitive ability, although not having the capability of converting his/her brain command to effective movements (an example is the case of amyotrophic lateral sclerosis – ALS, in its deepest degree).
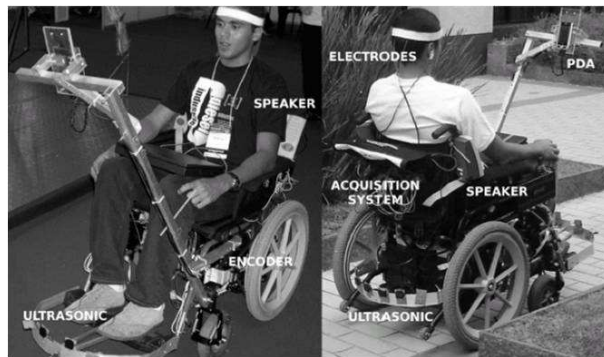


Figure 1: The robotic wheelchair being developed at UFES.

In this context, several research groups have proposed methods for preprocessing, feature extraction and classification of EEG patterns for BCI usage. Kim et al. (2003) perform an offline classification of motor tasks using 12 s epochs for the EEG channels C3 and C4 (considering the international 10-20 system of EEG electrodes distribution (Böcker et al., 1994)), filtering the signal for isolating the $\alpha$ (8-12 Hz) and $\beta$ (16-25 Hz) frequency bands and using the coefficients of Autoregressive Model (ARM) as the inputs for a linear discriminant classifier. Jia et al. (2004) perform an offline classification of motor imagery tasks using 9 s epochs for the EEG channels C3 and C4, filtering the signal for the 10-12 Hz passband and using Linear Discriminant Analysis (LDA) for classification. Liu et al. (2005) perform an offline classification of finger movements based on premovement potencial (Bereitschaftspotential) using the EEG channels C3 and C4, filtering the signal into two distinct frequency bands (0-3 Hz and 9-31 Hz), and using Common Spatial Subspace Decomposition (CSSD) and Artificial Neural Network (ANN) for classification. Anderson et al. (2007) perform a pseudo-online classification of three mental tasks using 32 electrodes

and filtering the signal in a passband of 8 to 30 Hz. Short-Time Principal Component Analysis (STPCA) is used for feature extraction and LDA is used for classification. Finally, Blankertz et al. (2007) perform an online classification of pre-movement potentials with user feedback, using 128 electrodes, Common Spatial Pattern (CSP) and LDA.

For an online approach, parameters such as signal sampling rate, size of time windows and their overlapping percentage should be determined (the above mentioned BCIs propose empirical values for such parameters). In our studies, an analysis of Auto-correlation Function (ACF) is performed to estimate the best values of such parameters. Power Spectral Density (PSD) is used for feature extraction, and a Bayesian classifier is used to recognize the mental tasks.

For each specific mental task performed, different preprocessing techniques are used. Therefore, a prior knowledge of the physiology involved in the task accomplishment influences the classification results. Even studying the physiological effects of a mental task in a general population, isolated individuals deviate from the average characteristics (Pfurtscheller et al., 1999). Then, it would be necessary an individual physiological study in order to maximize the performance of the classifier. In such context, this paper uses the Kullback-Leibler (KL) symmetric divergence as a method to standardize the selection of the positions of the electrodes and sampling frequency, in order to automatically adapt the BCI to motor or non-motor mental tasks.

## 2 MATERIALS AND METHODS

This work uses EEG data provided by IDIAP Research Institute (Switzerland), for the BCI Competition III, in June 2005 (Millán, 2004.). This data set contains EEG signals acquired from normal subjects during four sessions, without feedback. All 4 sessions of a given subject are 4 minutes long. In each session the subject randomly accomplished three mental tasks for about 15 seconds, which are the imagination of right or left hand movements and generation of words beginning with the same random letter. The EEG records are not split in trials, for the individuals are continuously performing one of the mental tasks. The acquisition system is a Biosemi one, using a cap with 32 integrated electrodes located at standard positions of the international 10-20 system. The sampling rate was 512 Hz. Signals were acquired at full DC, and no artifact rejection or correction was employed.

### 2.1 EEG Properties During Mental Tasks

The initial cognitive activity responsible for the intention of performing a motor task, either limb movement imagination or the physical movement, occurs in the brain cortex, over the

frontal, pre-frontal and parietal lobes. This signal is propagated to the striatum and the motor loop at the base of the brain and reaches the motor cortex through the thalamus. In the motor loop, this stimulus suppresses the sending of pacemaker rhythms in the thalamus to the primary motor cortex (M1), and starts the sending of signals related to the motor tasks performing (Bear et al., 2008).

Similarly, studies of Pfurtscheller et al. (1999) show that the accomplishment of mental tasks causes changes in the energy of $\alpha$ and $\beta$ frequency bands, with respect to the values observed in the cortex at rest. If such change is a decrease of the EEG power in such frequency bands, it is called an Event-Related Desynchronization (ERD). In opposition, an increase is called an Event-Related Synchronization (ERS). The ERD/ERS is observed in the cerebral hemisphere with opposite laterality in relation to the member imagined, due to the contra laterality of motor movements, since the signal path from M1 to the effective muscles, called pyramidal tract, is crossed. All axons of the pyramidal tract neurons intersect, either in the decussation of the bulb pyramids or in spinal cord (Guyton and Hall, 2006).

Bianchi et al. (1998) used the analysis of ERD/ERS and the phase of the Power Spectral Density (PSD) of the filtered EEG signal in the $\alpha$ and $\beta$ bands, and confirmed the activation flow from the frontal area to the central area during the preparation of a movement. Therefore, the electrodes over the motor cortex, the frontal lobe and the parietal lobe, have information from the mental task that occurs during motor activation flow (Figure 2 - Left).

Related to the mental task of generating words beginning with the same random letter, it is known that language processing, understanding and speech production occurs in the Broca's area, while the combination of information and interpretation occurs in Wernicke's area (Figure 2 - Right). As a consequence, information about this mental task is measured by electrodes over the left hemisphere of the frontal and parietal lobes.
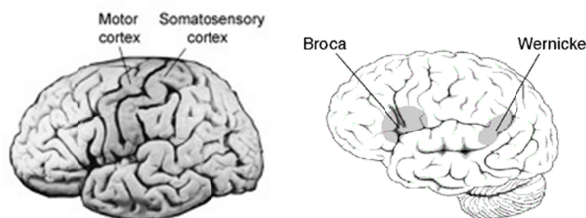


Figure 2: Left: Primary motor cortex; Right: Broca's area and Wernicke's area.

## 2.2 Signal Pre-Processing

Small time windows of EEG signals, with a fixed number of samples, $s$, were taken to simulate real-time classification. The time windows are continuously displaced by a sample (the sliding window technique). Thus, after the first time window is filled, each following window is generated by displacing the current window by one sample, preserving an overlapping of $s$-1 samples (Figure 3), so that the BCI classification rate is equal to the sampling rate.
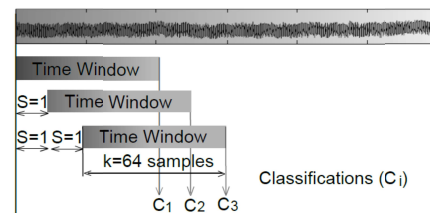


Figure 3: Sliding window technique.

Concerning the pseudo-online classification, it is necessary to estimate the optimal sampling rate, to not overload the system, and the optimal size of the time windows. EEG signals were originally obtained at a sampling frequency of 512 Hz, as previously mentioned. The analysis of the Autocorrelation Function, $R(\tau)$, was performed to determine if the signal is sub-sampled or over-sampled. The ACF is defined as

$$R(\tau) = E\left[\left(S(t) - \overline{S(t)}\right) \cdot \left(S(t - \tau) - \overline{S(t)}\right)\right],$$
$$\tau = 0, 1, \cdots, N \tag{1}$$

where E[•] is the expectation and the line above indicates average value along time. The nonlinear Auto-correlation Function is the ACF of the squared signal ($S^2(t)$). The sampling period, $T_S$, can be chosen by following the heuristic rule (Aguirre, 1995)

$$\frac{\tau_m}{20} \leq T_S \leq \frac{\tau_m}{10} \tag{2}$$

where $\tau_m = min(t_R)$, $t_R$ being the time instant correspondent to the first minimum of $R(\tau)$. Figure 4 shows the ACF of 9 channels. Horizontal lines represent the 95% confidence interval. The first minimum out of the confidence interval was obtained for a delay of 0.23 s ($\tau_m$), so that 0.011 s $\leq T_S$ $\leq$ 0.023 s, or 44.524 Hz $\leq F_S \leq$ 89.2857 Hz. As a result, the signal was re-sampled to 64 Hz, which corresponds to a three-step decimation.
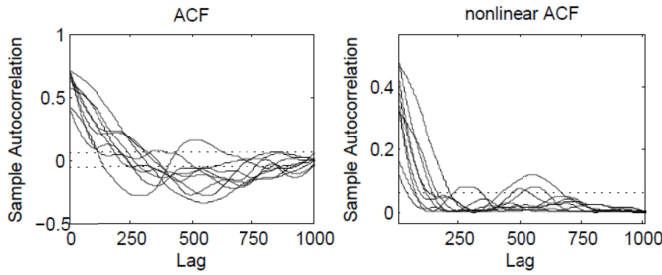
Figure 4: Autocorrelation Function.

The size of the time window should be sufficient to contain most of the influence of a sample in the next sample and to characterize a pattern that can be recognized by the classifier. To estimate the size of the time windows the coefficients of the Partial Auto-Correlation Function (PACF) were calculated for the EEG signals, to verify how $S(t)$ is autocorrelated to $S(t\text{-}\tau)$. The last lag out of Bartlett's confidence interval is the estimate of the minimum number of samples of the time windows (Lin et al., 1995). The PACF was performed in time windows of 10 s overlapped by 75%, for all EEG data. Figure 5 shows the mean of the PACF for all time windows and channels: a size of 5 s was obtained for individuals 1 and 3, and a size of 5.4 s was obtained for subject 2. Therefore, time windows with duration of 5 s were adopted.
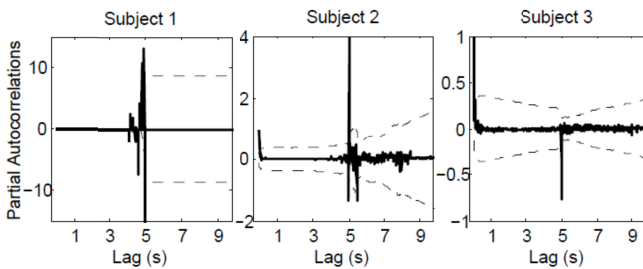


Figure 5: Partial Auto-Correlation Function.

## 2.3 Feature Extraction

For each time window, the power spectral density (PSD) of the EEG signals was calculated. PSD is the Fourier transform of the Auto-Correlation Function, $R(\tau)$, of the signal, if it can be considered Wide-Sense Stationary (WSS). It describes how the signal power is distributed along frequency, and is defined as

$$S(f) = \int_{-\infty}^{\infty} R(\tau) \cdot e^{-2\pi i f \tau} d\tau = F[R(\tau)] \qquad (3)$$

As the EEG signal was sub-sampled to 64 Hz the signal maximum frequency is 32 Hz. PSD was designed to return one coefficient for each integer value of frequency, thus resulting in 33 coefficients, including the DC component.

To view the effect of PSD in the class separation, data was clustered in each class and looking for a low-dimensional representation. K-means algorithm was used to cluster the data and obtain 20 centroids per class (Duda et al., 2000). Then Sammon mapping is performed for low-dimensional visualization (Sammon, 1969). Figure 6 shows the separation of classes of the subject-1 training data. Left figure shows the unprocessed EEG signal and right figure shows the signal processed through using PSD, where there is a quite visible class separation.
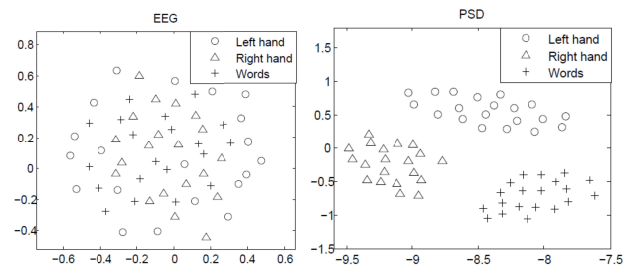


Figure 6: Left: Low-dimensional visualization of the class separation; Right: Low-dimensional visualization of the class separation using the PSD coefficients of the EEG signals.

## 2.4 Stationarity Test

As the PSD and the classifier require stationarity, the Runs Test was performed to determine the randomness of the EEG signal for distinct lengths of time windows. Figure 7 shows the mean of the z-score of the Runs Test calculated for time windows of 0.1s to 10 s. This z-score is compared with the z-score of the standard normal distribution at 5% significance level, which is 1.96. A z-score with an absolute value greater than 1.96 indicates non-randomness.
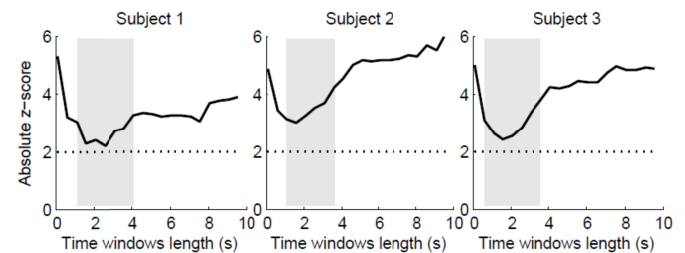


Figure 7: Runs Test z-score of the EEG signal for distinct time windows.

Figure 7 shows that the EEG is never considered random, since the z-score was always greater than 1.96. Nevertheless, the shaded area highlights the minimum z-score that corresponds to a greater degree of randomness. For subject 1 the EEG is more likely to be stationary for windows with lentgh under 1 s or above 4 s. For subjects 2 and 3 the EEG is more likely to be stationary for windows with lentgh under 0.5 s or above 3.6 s. As the PACF analysis obtained time windows with duration of 5 s, we can assume that the EEG is stationary.

## 2.5 Feature Selection

The aim of the feature selection procedure is to maximize the distinction between classes (mental tasks) in order to make the classification easier. Features used in the classifier are the PSD coefficients, and then the histograms of each frequency of the PSD spectrum are calculated for each EEG channel, resulting in 32x32 histograms for each class, discarding the DC component. Then, a measure of the discrimination between classes, D, was used to compare the three class histograms of each cell of the Channels X Frequency matrix. Figure 8 shows the Channels X Frequency matrix, where $C_1$, $C_2$ and $C_3$ are the three classes of mental tasks aforementioned.
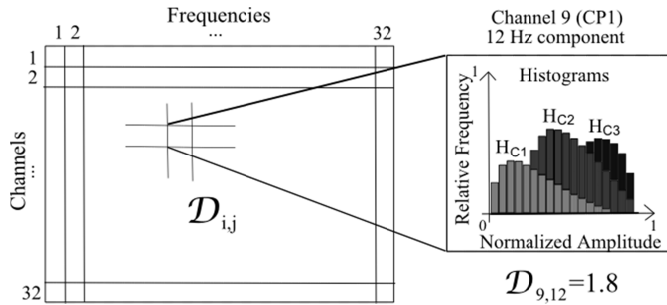


Figure 8: Channels x Frequencies matrix.

The Kullback-Leibler symmetric divergence was used to calculate the discrimination measure $D$ (Rossow et al., 2010). The K-L divergence is given by

$$D_{KL}\left(C_1 || C_2\right) = \sum_{i=1}^{n} H_1\left(i\right) \log \left( \frac{\max\left(H_1\left(i\right), \delta\right)}{\max\left(H_2\left(i\right), \delta\right)} \right) \quad (4)$$

where $H_1$ and $H_2$ are histograms of two distinct classes, $C_1$ and $C_2$, with n samples and $\delta$ is a small number, here selected as $\delta = 0.001$, used to prevent the logarithm of zero or the division by zero. The symmetric version of the K-L divergence is given by

$$S\left(C_1 || C_2\right) = D_{KL}\left(C_1 || C_2\right) + D_{KL}\left(C_2 || C_1\right) \quad (5)$$

Then, the discriminant $D_{i,j}$ of each channel $i$ and frequency $j$ is given by

$$D_{i,j} = S_{i,j}\left(C_1 || C_2\right) + S_{i,j}\left(C_1 || C_3\right) + S_{i,j}\left(C_2 || C_3\right) \quad (6)$$

Figure 9 shows the discriminant $D_{i,j}$ Matrix calculated for subject 2, in which the discriminant values are represented in grayscale. Some highest discriminant values were obtained for the tuples (CP1, at 12 and 22 Hz) and (C4, at 12 and 21 Hz), which is conform to prior knowledge that motor mental tasks affect $\alpha$ and $\beta$ frequency bands over the contralateral motor cortex. The tuples (F3, at 32 Hz) and (Fz, at 28 Hz) also obtained high discriminant values, what we believe to be related to mental task of generation of words that would be occurring in Broca's area, which is in the left brain hemisphere.

All features having discriminant values over a percentage, $\xi$, of the maximum discriminant obtained, which in this case was obtained for channel CP1 at 12 Hz, were chosen. Then, figure 10 shows the selected features of subject 2 discriminant matrix, using $\xi = 50\%$.

## 2.6 Classifier

The Bayesian classifier is a simple probabilistic approach based on Bayes' Rule in which the class distributions are modeled by a normal distribution (Duda et al., 2000). The posteriori probability that the correct class is $k$, given a data sample $x$, is

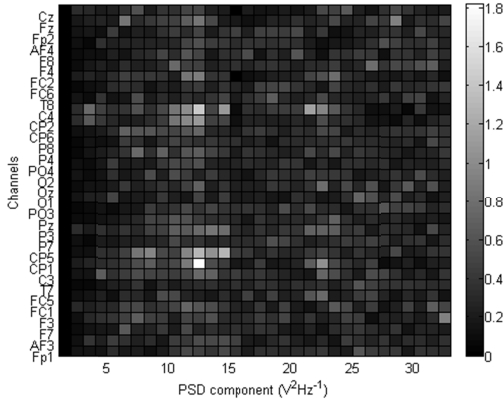$$P\left(C_k | x\right) = \frac{P\left(x | C_k\right) \cdot P\left(C_k\right)}{P\left(x\right)} \quad (7)$$

Figure 9: Discriminant matrix, $D$, of Subject 2.

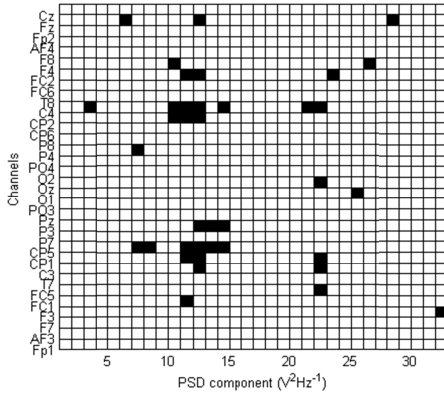

Figure 10: Selected features using $\xi$=50%.

where P$(x|C_k)$ is the class probability density function (pdf), $P(C_k)$ is the priori probability and P$(x)$ is the unconditioned probability. The normal distributions of each class $k$ are modeled using the same covariance matrix, $\Sigma$, and are given by

$$P\left(x|C_k\right) = \frac{1}{(2\pi)^{\frac{d}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^T\Sigma^{-1}(\mathbf{x}-\mu_k)} \quad (8)$$

where $\mu_k$ is the mean of the class $k$ and $d$ is the dimension of the normal distribution. The linear discriminant function, $f_k$, is given by the logarithm of the posteriori probability, as in (6) . The unconditioned probability, P$(x)$, is only a scale factor for the posteriori probability, $P(C_k|x)$, so that the term -$ln$P$(x)$ can be ignored. If all classes have equal priori probabilities, $P(C_k)$, the term $lnP(C_k)$ can also be ignored. For the linear case of the Discriminant Analysis, it is

considered that all classes have the same covariance matrix, so $(-1/2)ln|\Sigma|$ and $(-d/2)ln2\pi$ are independent of $k$. They become unimportant additive constants that can be ignored. The expansion of the quadratic form $(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)$ results in a sum involving $x^T\Sigma^{-1}x$ that is also independent of $k$ and can be ignored (Duda et al., 2000), resulting in

$$f_k(x) = \ln\left(P(C_k|x)\right) \Rightarrow$$

$$f_k(x) = \ln\left(P\left(x|C_k\right)\right) + \ln\left(P(C_k)\right) - \ln\left(P\left(x\right)\right) \Rightarrow$$

$$f_k(x) = -\tfrac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k) - \tfrac{d}{2}\ln 2\pi - \tfrac{1}{2}\ln|\Sigma| \Rightarrow$$

$$f_k(x) = -\tfrac{1}{2}x^T\Sigma^{-1}x + \mu_k^T\Sigma^{-1}x - \tfrac{1}{2}\mu_k^T\Sigma^{-1}\mu_k \Rightarrow$$

$$f_k(x) = \mu_k^T\Sigma^{-1}x - \tfrac{1}{2}\mu_k^T\Sigma^{-1}\mu_k$$

$$(9)$$

The class of a data sample $x$ is assigned to the class with higher linear discriminant function value. Two sessions of the database were used to estimate the means and covariance matrices of each class, i.e. 66.67% of the database. Then, the covariance matrix used by the classifier is the mean of the covariance matrices of the three classes. One session was used for validation, and all events are equally likely a priori.

## 2.7 Reclassification Method

Figure 11 shows the classifier output. The best value of success rate of subject 1 was 83.77±1.27%, with $\xi$ = 50%. For subject 2, the best value of success rate was 65.75±1.52%, with $\xi$ = 50%, and the best value of success rate of subject 3 was 55.53±1.59%, using $\xi$ = 0%.

Real class is represented by dotted line and the predicted class is represented by the continuous line. In this figure, Class 1 corresponds to a task of imagination of left hand movement; Class 2 corresponds to the same task for the right hand and Class 3 corresponds to the mental task of generating words beginning with the same random letter.

It may be noted that the predicted class oscillates faster than the user could change tasks, which represents instability of the classifier. Therefore, a method to smooth the classifications was developed. Each classifier output will form a new vector of classification, which stores the previous classifications for a period of time, called reclassification window. These windows were taken to behave as time windows; they are continuously displaced by one classification, thus, not changing the original classification rate of the system.

The reclassification could smooth the results using the most frequent class in the reclassification window, but this will lead to a delay during the transitions of mental tasks, since the new class should occur more often than the old class so

that the classifier identifies a change of class. To minimize the delay, the output of the classifier will be the class with highest weight in the reclassification window. The weight of each past classification is given in relation to the size of the subgroup of equal classes that it belongs. The weight is calculated by the inverse of the probability of a repeated occurrence of a class. For three equiprobable classes, a sample $s$ in a subgroup with size $n$ will have probability $P(s) = (1/3)^n$. Thus, the weight of this sample is given by $W(s) = (3)^n$, penalizing, this way, isolated samples or samples in small subgroups among the various identical classifications, which would be a noise in the classification. Figure 12 shows an example where the reclassification window is composed of nine past classifications. Although the 3 classes occur with the same number of times, the class that occurs more times without changing will have the highest weight and will be assigned to the system output.

## 3 RESULTS AND DISCUSSION

The length of the reclassification windows, $\lambda$, was varied from 1 to 320 past classifications. The values of $\xi$ are varied up to 90% in order to find the best parameter for each subject. Figure 13 shows the curves obtained for each subject, where the circle mark indicates the best values of $\xi$. Figure 14 shows the stabilized classification of the signal using the values of $\xi$ and the reclassification windows size, $\lambda$, that obtained the best result (it can be noticed a significant reduction in the noise during the classification). Table 1 shows the best values of success rate of each subject and the parameters ($\xi$, $\lambda$) used. In the last column, B is the information transfer rate, calculated, in [bits/min], as $(60/\lambda) \bullet B$, with

$$B = \log_2 N + p \cdot \log_2 p + (1-p) \cdot \log_2 \left( \frac{1-p}{N-1} \right), \quad (10)$$

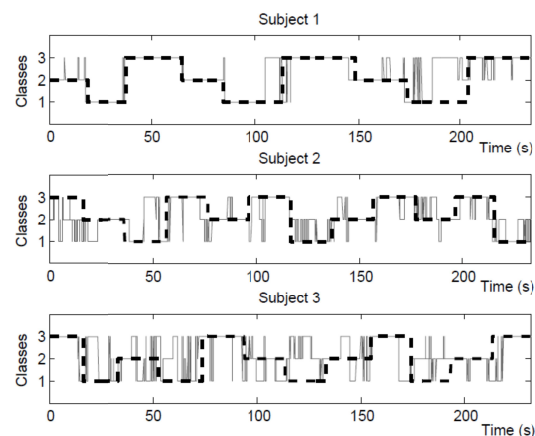where $N$ is the number of classes and $p$ is the success probability.
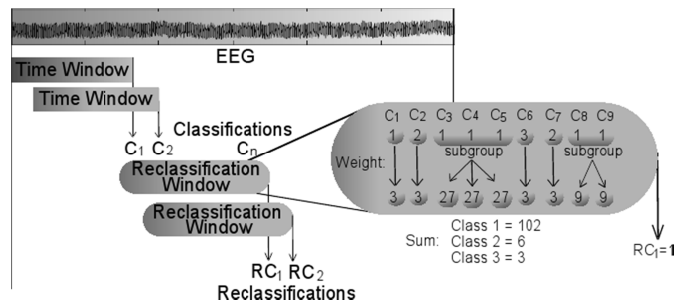


Figure 11: Classifier output.



Figure 12: Time windows and the reclassification windows.
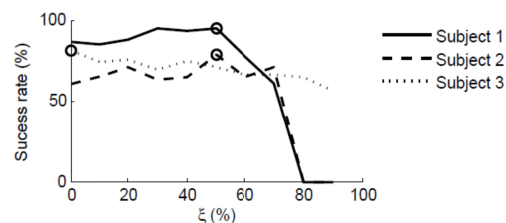


Figure 13: Classifier sucess rate for different choices of $\xi$.

Table 1:

| | Accuracy | $\xi$ | $\lambda$ | B [bits/min] |
|---|---|---|---|---|
| Subject 1 | 94.93±0.71% | 50% | 2.5s | 29.88 |
| Subject 2 | 78.89±1.31% | 50% | 1.88s | 20.12 |
| Subject 3 | 81.31±1.25% | 0% | 5s | 8.44 |

The Table 2 shows the statistical measures of the performance of the best classifiers, where it is shown the Precision, Sensitivity and Specificity of each class ($C_1$, $C_2$ and $C_3$).
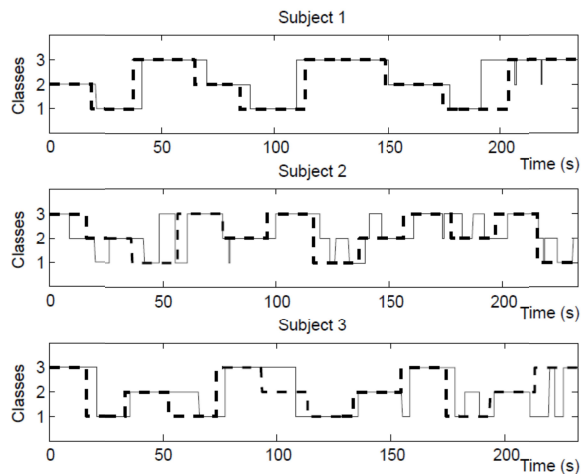


Figure 14: Stabilized classifications.

Table 2:

| | Precision | Sensitivity | Specificity |
|---|---|---|---|
| **Subject 1** | | | |
| $C_1$ | 0.85 | 1 | 0.95 |
| $C_2$ | 0.98 | 0.97 | 0.99 |
| $C_3$ | 0.99 | 0.91 | 0.99 |
| **Subject 2** | | | |
| $C_1$ | 0.54 | 0.79 | 0.87 |
| $C_2$ | 0.80 | 0.73 | 0.86 |
| $C_3$ | 0.93 | 0.85 | 0.95 |
| **Subject 3** | | | |
| $C_1$ | 0.80 | 0.79 | 0.90 |
| $C_2$ | 0.80 | 0.79 | 0.91 |
| $C_3$ | 0.83 | 0.86 | 0.91 |

Finally, the results presented here are compared with the best results submitted to BCI Competition III. Those works used the same data set used here, so that they have the same mental tasks for classification. The results are summarized in Table 3, obtained from Anderson et al. (2007).

First, some considerations should be made. The classifications performed by the works compared in Table III used a simple reclassification method, smoothing the classifications results with windows of 0.5 seconds, which causes a highly probable delay during the classification. As one can see, our results are in average 23% better than the best work sum-

Table 3:

| | Subject 1 | Subject 2 | Subject 3 | Average |
|---|---|---|---|---|
| Authors | 94.93 | 78.89 | 81.31 | 85.04 |
| C. Anderson | 62.3 | 57.6 | 47.5 | 55.8 |
| S. Sun | 74. 3 | 62.3 | 52.0 | 62.8 |
| A. Schlögl | 69.0 | 57.1 | 32.3 | 52.7 |
| E. Arbabi | 55.4 | 51.8 | 43.6 | 50.2 |
| A. Salehi | 26.5 | 32.8 | 24.5 | 28.0 |

marized in Table III, with reclassification windows of up to 5 seconds being used. However, the reclassification method developed causes less likely delays, which partly justifies the use of larger reclassification windows.

It should also be taken into account the different technique adopted in each compared work. Anderson et al. used Short-time PCA for feature extraction and LDA for classification. Sun et al. removed artifacts from seven electrodes, the EEG signal was bandpass filtered to 8-13 Hz for subject 1 and 2 andto 11-15 Hz to subject 3. A multiclass approach to common spatial patterns was used and support vector machines (SVM) was used for classification. Schlögl et al. downsampled the signal to 128 Hz, formed all bipolar channels and estimated autoregressive models for each channel, besides getting the energy in $\alpha$ and $\beta$ bands. Then, the best single feature was selected to a statiscal classsification. Arabi et al. downsampled the signal to 128 Hz, filtered it preserving the 0.5-45 Hz band, extracted features based on statistical measures and on parametric models of one second windows and classified them using a Bayesian classifier. Salehi used a combination of Short-time Fourier transform (STFT) energy values and time-domain features to be classified by a Bayesian classifier.

Finally, the work presented here downsampled the EEG to 64 Hz, used the PSD with five second windows for feature extraction, the symmetric KL divergence for feature selection and a simplified version of a Bayesian classifier.

## 4 CONCLUSION

The application of all proposed methods resulted in a classifier able to identify the three classes and to obtain results much above random (33%) success rate. Finally, we conclude that the success rate, the stability of the classifier result and the information transfer rate make possible its application in BCIs associated to robotic devices, such as a robotic wheelchair, whose control should be performed in real time.

As future work we intend to study the use of the multivariate gamma distribution to describe the PSD features, instead of the multivariate normal distribution used in the classifier. We also suggest the use of different types of classifiers associ-

ated with the methods described in this study, like SVMs or Regularized Discriminant Analysis (RDA).

## ACKNOWLEDGMENT

## REFERENCES

Aguirre, L. A (1995). "A Nonlinear Correlation Function for Selecting the Delay Time in Dynamical Reconstructions". *Physics Letters A*. Vol. 203, pp. 88-94.

Anderson, C. W; Knight, J. N; Kirby, J. M and Hundley, D. G (2007). "Classification of Time-Embedded EEG using short-time principal component analysis". In: Dornhege, G; Millán, J. M; Hinterberger, T; McFarland, D. J and Müller, K. R., eds. 2007. *Toward Brain-Computer Interfacing*. Cambridge: The MIT Press, pp.261-278.

Bear, M. F.; Connors, B. W. and Paradiso, M. A (2008). *Neurociências: Desvendando o Sistema Nervoso*. 3ª ed. São Paulo: Artmed® S. A..

Bianchi, A. M; Leocani, L; Mainardi, L. T; Comi, G and Cerutti, S (1998). "Time-Frequency Analysis of Event-Related Brain Potencials". *Proceedings of the 20th Annual Internacional Conference of the IEEE Engineering in Medicine and Biology Society*. Vol 20, no.3, pp.1486-1489.

Blankertz, B; Dornhege G; Lemm S; Krauledat, M; Curio, G e Müller, K. R (2007). "The Berlim Brain-Computer Interface: Machine Learning Based Detection of User Specific Brain States". In: Dornhege, G; Millán, J. M; Hinterberger, T; McFarland, D. J and Müller, K. R., eds. 2007. *Toward Brain-Computer Interfacing*. Cambridge: The MIT Press, pp.85-102.

Böcker, K. B. E., Avermaete, van, J. A. G. and Berglenssen, van den, M. M. C (1994). "The International 10-20 System Revised: Cartesian ans Spherical Coordinates". Brain Topogr. Vol. 6, pp. 231-235.

Duda, R. O; Hart, P. E and Stork, D. G (2000). *Pattern Classification*, 2rd ed. United States: John Wiley & Sons Inc., chapter 10: Unsupervised Learning and Clustering, pp.13-16, chapter 4: Nonparametric Techniques, pp.18-36, and chapter 2: Bayesian Decision Theory, pp.3-29

Guyton, A. C. and Hall, J. E (2006). *Textbook of Medical Physiology*. 11ª Ed. Philadelphia: Editora.

Jia, W; Zhao, X; Liu, H; Gao, X; Gao, S and Yang, F (2004). "Classification of Single Trial EEG during Motor Imagery based on ERD". *Proceedings of the 26th Annual Internacional Conference of the IEEE Engineering in Medicine and Biology Society*, pp.5-8.

Kim, J. A; Hwang, D. U; Cho, S. Y; Han, S. K (2003). "Single Trial Discrimination between Right and Left Hand Movement with EEGSignal". *Proceedings of the 25th Annual Internacional Conference of the IEEE Engineering in Medicine and Biology Society*, pp.3321-3324.

Lin, F; Huo, X. Y; Gregor, S and Irons, R (1995). "Times Series Forecasting with Neural Networks". *Complexity Internacional*, Vol. 2.

Liu, B; Wang, M; Li, T and Liu, Z (2005). "Identification and Classification for finger movement based on EEG". *Proceedings of the 27th Annual Internacional Conference of the IEEE Engineering in Medicine and Biology Society*, pp.5408-5411.

Millán, J.del R (2004). "On the Need for On-line Learning in Brain-Computer Interfaces". Proceedings of the International Joint Conference on Neural Networks, Hungary, pp.2877-2882.

Pfurtscheller, G. and Lopes da Silva, F. H (1999). "Event-Related EEG/MEG Synchronization and Desynchronization: Basic Principles". Elsevier, Clinical Neurophysiology, Vol. 110, pp. 1842-1857.

Rossow, A. B; Salles, E. O. T. and Côco, K. F (2010). "Classificação Automática de Estágios do Sono pela Análise do Sinal de EEG Através de Wavelet Packet". *Anais do XVIII Congresso Brasileiro de Automática*, pp.3825-3832.

Sammon, J. W (1969). "A Nonlinear Mapping for Data Structure Analysis". *IEEE Transactions on Computers*, C-18 (5) , pp. 401-409.