FEATURE EXTRACTION EVIDENCING LONG TERM DEPENDENCIES FOR A BRAIN-COMPUTER INTERFACE

A. M. Lazar∗ and R. Ursulean∗∗

∗University of Medicine and Pharmacy, Bioengineering, Iasi, Romania
∗∗ Technical University, Electrotechnics and Electrical Machines, Iasi, Romania

anca@scs.etc.tuiasi.ro
ursulean@ee.tuiasi.ro

Abstract: The electroencephalographic signals (EEG), in addition to their clinical applications, can be used as support for direct brain-computer communication devices (Brain-Computer Interfaces BCIs). During mental imagination of specific movements, EEG patterns that characterize them emerge. Signals, recorded from sensorimotor areas of subjects with severe motor disabilities trained to use their mu rhythms (8-12 Hz) for cursor control, are exploited. The obtained EEG signals may be characterized as statistical self-similar and a measure for this is their fractal dimension, which may be computed by means of the Hurst coefficient. The Hurst coefficient is also used to choose the regularity of the wavelet involved in the BCI. Hence, provided that the number of vanishing moments of the chosen wavelet is sufficiently large the correlations will decay rapidly. A method using the continuous wavelet transform and Student t-statistic was proposed for feature extraction of mu rhythms. The complex Morlet wavelet was chosen as it has a good time-frequency resolution. The results suggest that computing the local extrema of the t-value scalogram of two groups may represent a good choice for feature detection of mu rhythms.

Introduction

A brain–computer interface (BCI) is a system for direct communication between brain and computer, exploiting the electrical activity of the brain for those persons incapable of motor functions, but having cognitive abilities. A BCI attempts to express the user's intent by using signal processing and pattern recognition techniques to translate control signals into reliable device commands.

The input signals may be either scalp electroencephalogram (EEG) or cortical potential which provides higher signal to noise ratio [1].

Dealing with EEG signals analysis is not an easy task when one has to decide upon a strategy to develop a brain computer interface. There are quite a few characteristics of the signals representing the brain activities that make usual strategies to fail: 1) the signals are nonstationary and therefore tools like the Fourier transform are cumbersome, 2) some of them have obviously chaotic behaviour and 3) the latter contain statistical self-similarities (i.e. they roughly look the same – any section of the data set would have the same statistical properties as any other).

Another feature of the EEG signals is their behaviour as long-memory processes. In order to detect such a conduct, a reasonable option is to compute the Hurst coefficient. This coefficient, first introduced by H. Hurst, is a measure of the statistical self-similarity and long-memory process [2]. Performing such an analysis is useful to detect the long memory dependencies (long-term autocorrelations), which establish undesirable effects on amplitude distribution of the data. Knowledge of the Hurst coefficient may be a good criterion to choose a special kind of wavelet when performing the "whitening" of a time series by means of the wavelet transform [3].

A BCI can be implemented either using the responses of the brain to stimuli [4] (e.g. P300 potentials) or by training the user to control his/hers brain waves [5], [6] (e.g. self-regulated mu and/or beta rhythms).

In this paper, we focus our attention on mu rhythms produced by imagination of moving a cursor on a screen of a monitor in order to reach a target located on the top or bottom edge of the screen. The mu rhythm, defined as being a part of alpha rhythm, is an activity generated in sensorimotor cortex and recorded over central head regions in the 8-12 Hz frequency band. The mu rhythm decreases or desynchronizes with the movement or even only with the imagination of movement [7]. This rhythm was chosen because it is produced in those places which are directly related to movement and because it was demonstrated that a person, disabled or not, is capable to be trained in order to control the amplitude of the mu rhythm. BCIs using information extracted from mu and beta rhythms are considered independent BCIs; their operation does not depend in any way on the brain's normal output pathway. The increases/decreases of this rhythm have already been used several times as a support for a BCI [1], [7].

Effective feature characterization and extraction is essential for the development of a BCI. Feature extraction can be performed in different ways by means

ISSN: 1727-1983 © 2005 IFMBE

of temporal, spectral, spatial or wavelet analysis, sometimes using statistical properties of the EEG recordings.

We propose to exploit the high redundancy of the continuous wavelet transform (CWT) for precise time-space (time-frequency) localization of the mu rhythm and to perform the Student's t-test applied to the CWT-s of two independent samples in order to find the local maxima to discriminate between the movement of the cursor to the top or bottom of the right edge of the screen.

Evidencing long-term dependencies by means of the Hurst coefficient is presented in what it follows.

The complexity of self-similar structures is quantified by their usually non-integer fractal dimension [8]. Between the fractal dimension *D*, the topological dimension *T* (T = 1 for a time series) and the power law exponent γ of the spectral density, also known as "self-similarity parameter" [9],

$$S(f) \propto \frac{1}{f^{\gamma}} \tag{1}$$

there is a simple relationship:

$$D = T + \frac{3-\gamma}{2}. \tag{2}$$

Another one that links the fractal dimension and the Hurst coefficient is given by

$$D = T + 1 - H. \tag{3}$$

Therefore, knowing the Hurst coefficient is a possible approach to determine the fractal dimension of the EEG signals and the power law exponent.

Let X_t be a stochastic process whose behaviour is known at the discrete time points $t \in \{0,\ldots,N\}$ and *n* a small number relatively to *N*. Let us denote by *C* the integer part of *N/n*

$$C = \left[\frac{N}{n}\right]. \tag{4}$$

In other words, *C* is the number of subintervals of length *n*, obtained from the initial set of points.

The limits of each subinterval are therefore given by

$$(c-1)n \le t \le cn \text{ and } c \in \{1,\ldots,C\}. \tag{5}$$

For every subinterval the datum is "corrected" by means of the slope of the process for each subinterval,

$$X_t \leftarrow X_t - \frac{\left(X_{cn} - X_{(c-1)n}\right)t}{n} \tag{6}$$

and, in this way, the smallest box with sides parallel to the coordinate axes is build. The box has obviously vertical axes at $t = (c-1)n$ and $t = cn$ and the height given by

$$R_c = \max_{(c-1)n \le t \le cn}\left\{X_t - \frac{(X_{cn} - X_{(c-1)n})t}{n}\right\} - \min_{(c-1)n \le t \le cn}\left\{X_t - \frac{(X_{cn} - X_{(c-1)n})t}{n}\right\} \tag{7}$$

If we denote by S_c the empirical standard error, i.e.

$$S_c = \max_{(c-1)n+1 \le t \le cn}\left|X_t - X_{t-1}\right|, \tag{8}$$

Dividing R_c by S_c corrects for the scale inhomogeneity in the case of a nonstationary process, as an EEG signal proves to be.

The total area of the boxes, corrected for scale, depends on *n* and it is proportional to

$$\left(\frac{R}{S}\right)_n = \frac{1}{C}\sum_{j=1}^{C}\frac{R_j}{S_j}. \tag{9}$$

Choosing the suited wavelet transform is important for the feature extraction used in the implementation of a BCI.

Wavelets have been used in biomedical engineering since the end of the 1980's. The wavelet analysis uses a simple idea which concerns in expanding the signal on a set of dilated (compressed) and translated functions $\psi\left(\frac{t-b}{a}\right)$, where *a* is the scale and *b* the time shift (both of them measured in units of times). $\psi(t)$ is called *mother wavelet*. When proper wavelet tools are chosen, the wavelet analysis produces better results than Fourier transforms methods do.

First of all, we have to choose the type of the wavelet transform, the type of the mother wavelet and scales.

For the first step, the decision of the type of the wavelet transform, we must take into account that in this application we are faced to detect differences from two EEG signals with or without mu rhythms and to use wavelet transform as a kind of "*template matching*" [4]. For this purpose the redundancy becomes a quality, so the continuous wavelet transform is well suited to be preferred.

The continuous wavelet transform of a finite energy signal, *f(t)*, in terms of the two variable, *a* and *b* ($a \in (0,\infty)$ and $b \in R$) is defined as follows [10]

$$CWT(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}, \tag{10}$$

where $\overline{\psi}$ denotes the complex conjugate of ψ.

The wavelet type is set according the features need to be extracted from the signal *f(t)* (shape, length and smoothness). The time-bandwidth product of the wavelet transform is the square of the input signal and for most practical applications this is not a desirable property. Therefore one imposes some additional conditions on the wavelet functions in order to make the wavelet transform decrease quickly with decreasing scale. These are the *regularity conditions* and they state that the wavelet function should have some smoothness which is related to the number of *vanishing moments R*. Regularity is a quite complex concept and in [11] the author explains it using the concept of vanishing moments. The *moments* M_r of the wavelet are defined as

$$M_r = \int t^r \psi(t)\,dt . \tag{11}$$

The number of vanishing moments of a mother wavelet $\psi(t)$ is named to be the largest integer R that satisfies $M_r = 0,\ \ \overline{r = 0, R-1}$. The Hurst coefficient H is also used to choose the regularity of the wavelet. Hence, provided that the number of vanishing moments of the chosen wavelet is sufficiently large [3],

$$R > 2H + 1 , \tag{12}$$

the correlations will decay rapidly.

The last step, the option for scales, is linked to specific signal properties and based on how many details in what frequency range will be computed.

The main topic in analysis of mu rhythm is to get a time-scale localization of components to identify the difference between imagination of moving toward top or down of one edge of a screen. The significant difference in 8-12 Hz band is assessed by comparing the mean value of each of the wavelet transforms of the two groups by means of a two-sample Student's t-test (two independent samples to test for the null hypothesis that the mean values are equal [12]). If the null hypothesis is rejected, the statistically significant appearance of mu rhythms changes in time-frequency domain is confirmed.

We consider two samples (groups) of N_s trials each and k channels of EEG recordings.

The steps we have to follow in order to perform two-sample Students t-test are:

1) The $CWT^{kn}(a,b)$ of the signal $f_n^k(t)$ is computed for each channel k and each trial n according to (10).

2) The mean $\overline{CWT_s^k(a,b)}$ is computed for the two samples s as it follows

$$\overline{CWT_s^k(a,b)} = \frac{1}{N_s}\sum_{n=1}^{N_s} CWT^{kn}(a,b), \qquad s = \overline{1,2} \tag{13}$$

3) The variances $\sigma_s^k(a,b)$ are computed for the two samples s as follows

$$\sigma_s^k(a,b) = \frac{1}{N_s - 1}\sum_{n=1}^{N_s}\left(CWT^{kn}(a,b) - \overline{CWT^{kn}(a,b)}\right) \tag{14}$$

4) The two-sample t-statistic $t^k(a,b)$ is computed. The formula for the t-statistic is given below

$$t^k(a,b) = \sqrt{\frac{N_1 N_2}{N_1 + N_2}}\,\frac{\overline{CWT_1^k(a,b)} - \overline{CWT_2^k(a,b)}}{\sigma_p^k(a,b)} \tag{15}$$

where $\sigma_s^k(a,b)$ is the pooled standard deviation

$$\sigma_p^k(a,b) = \sqrt{\frac{\sigma_1^2(a,b)\cdot(N_1 - 1) - \sigma_2^2(a,b)\cdot(N_2 - 1)}{N_1 + N_2 + 1}} \tag{16}$$

of the two samples having the variances $\sigma_1^2(a,b)$, respectively $\sigma_2^2(a,b)$ (the channel number k was deliberately omitted for the sake of simplicity). The calculated t-statistic describes the difference between two samples means in standardized units.

From the probability table of the t-statistic [12], the t value associated with the degrees of freedom and the significance level is found. This value is called *the critical value of t*, denoted by t_{crit}. If our calculated values of *t* lies outside the probability interval for *t*, then there is less than 5% chance the two samples means to belong to the same population [12].

The local extrema (a_i^k, b_j^k) of the two-sample t-statistic $t^k(a,b)$ are found. At this step, we take into account a threshold equal to t_{crit}, this denotes the local extrema must be grater then t_{crit}. These points represent the local maxima difference between the two groups in the time-scale domain of the mu rhythms.

The $CWT^{kn_{i,j}}(a_i^k, b_j^k)$ (the continuous wavelet transform for each point (a_i^k, b_j^k) determined at the previous step) is computed and it represents the extracted feature for the two groups $s = \overline{1,2}$.

Materials and results

The data set was provided by the Wadsworth Center, New York State Department of Health [7] and represents records from three trained subjects (named AA, BB and CC), in 10 sessions (6 labelled and 4 unlabeled sessions), with 6 runs per session. The total number of trials per session was equally used for the four position of the target at the one edge of the screen: top, middle top, middle bottom and bottom. In our research, we used only those recording belonging to the cases when the subjects imagined moving the cursor to the top and bottom at the right site of the screen and they managed to reach the correct position. The EEG recordings were taken from 64 electrodes, each one referred to the right ear. All the signals were band-pass filtered (0.1-60 Hz) and sampled at 160 Hz sampling

ISSN: 1727-1983 © 2005 IFMBE

rate. The subjects used their mu rhythm amplitude to control movement of the cursor toward the position of a target located at the right edge of the screen. The detailed description of the experiment is found in [7].

Our objective was to use the first six labelled sessions to study the possibility of extracting representative features for BCI.

Because we are only interested in the activity of the motor cortex, only the electrodes around the sensorimotor cortex, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4, were chosen [13]. For all the three subjects, only two of the 15 chosen channels were used as the mu rhythm in these channels had the largest peak in power spectral densities and the significant difference between power spectral densities of the two positions of the target.

Data sets from all the subjects were used to compute the Hurst coefficient and the algorithm presented above was implemented. The changes for the coefficient were not significant from one subject to another, as can be seen in Table 1 in the case of the C3 and Cp3 signals. The second subject presented slightly higher values of the Hurst coefficient for the C3 signal.

Table 1: The Hurst coefficient and the fractal dimesion

Subject	Signal	Hurst coefficient H	Fractal dimension D
AA	C3	0.241	1.759
AA	Cp3	0.23	1.769
BB	C3	0.356	1.644
BB	Cp3	0.277	1.723
CC	C3	0.275	1.725
CC	Cp3	0.27	1.73

As a general observation, the values of the Hurst coefficient are significantly different from 0.5 which characterize a random signal. This is also a good indicator that "long memory" effects influence the signals.

The fractal dimension for the same task is also presented in Table 1. It can be seen that except for the BB subject in the case of signal C3 the fractal dimension of the recordings are different with less than 5%. This could be a measure of the accomplishment of the task when dealing with a "focused" demand.

The values of the fractal dimension of the time series for all the tasks led to the conclusion that the system that generated them has a manifest tendency towards a scale-free behaviour. Due to the long range correlations generated locally, the whole system operates in a critical state. This phenomenon was defined in [8] and [9] although for another type of system.

A complex Morlet wavelet is defined by

$$\Psi(x) = \frac{1}{\sqrt{\pi F_b}} \exp\left(2\pi j F_c x - \frac{x^2}{F_b}\right). \qquad (17)$$

It depends on two parameters: F_b is a bandwidth parameter, F_c is a wavelet center frequency.

As the complex Morlet wavelet no.1-1.5 has a good time-frequency resolution [4], [10] and the number of vanishing moments is $R=2$ we use it as the mother wavelet.

The continuous wavelet transforms $CWT^{kn}(a,b)$ and the means of modulus of complex-valued wavelet transform, $\overline{CWT_s^k(a,b)}$, of the groups were computed for each of the trials for the two groups. We have limited in the region of 8-12 Hz.

In figure 1 the t-value scalogram of mu rhythm for one of the three subjects was represented (channel CP3 and C3).

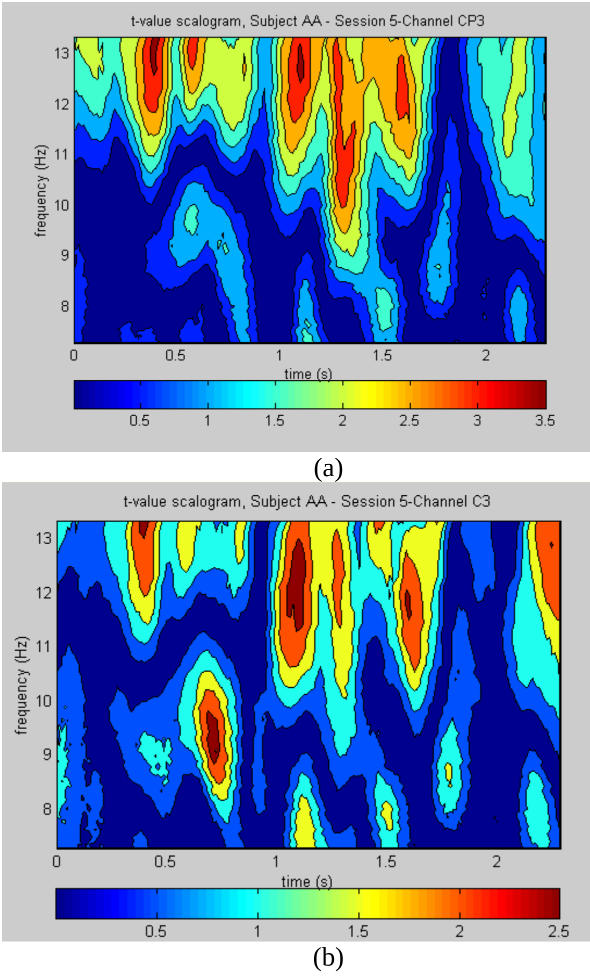
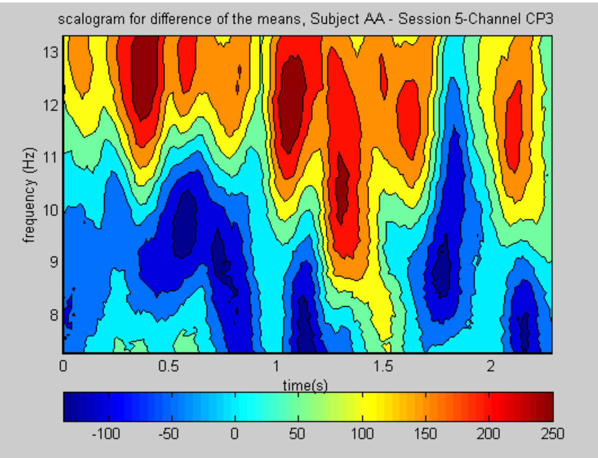


Figure 1: T-value scalograms of mu rhythm (a) CP3 channel and (b) C3 channel (subject AA)

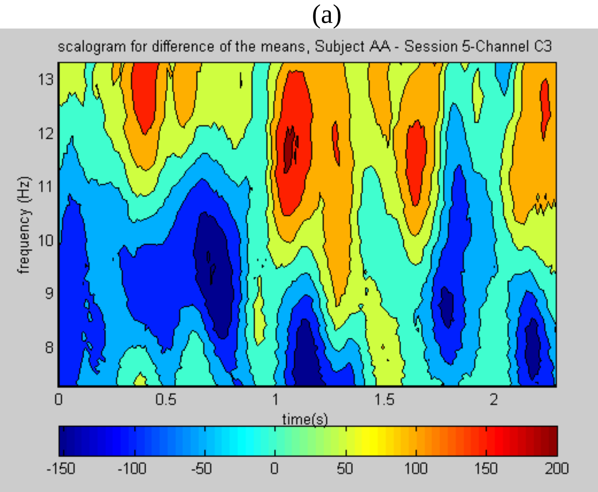
For CP3 channel, there is obvious that there are time domains where the t-value scalogram has high values (e.g. at 0.4-0.5 s and 1-1.2 s for the frequency domain around 11.5-13 Hz and at 1.25-1.35 s for the frequency domain nearly 9.5-13 Hz of the mu rhythm). As for C3 channel, the value are higher then for CP3 and the localization in time domain for frequencies of the mu rhythm is a little different, especially in 0.65-0.75 s time domain and 8.5-10 Hz in frequency domain.

ISSN: 1727-1983

Further, the scalogram of the difference average of the continuous wavelet transforms for top and bottom was displayed in figure 2. There are some distinctions between the t-value and difference average scalograms. So, for channel CP3, one late maximum at about 2,1 s in the difference average is not found by the t-value scalogram. For both channels, better time-frequency localization is acquired by the proposed method than by the difference average scalogram.



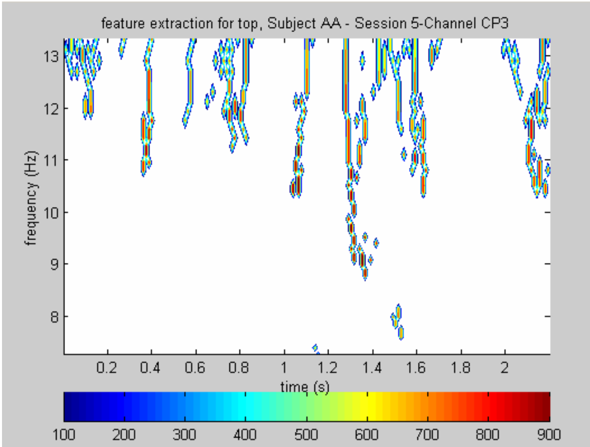
(a)



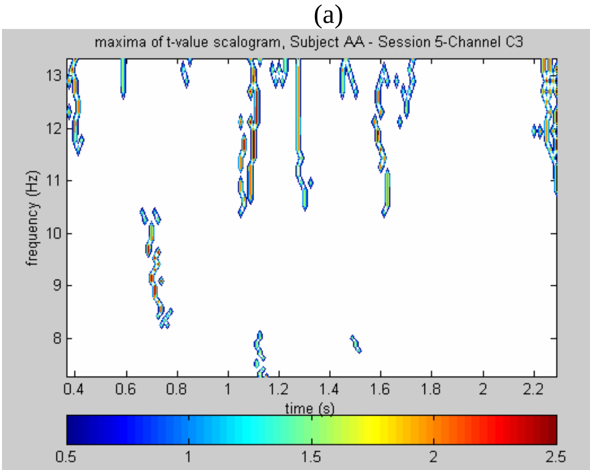
(b)

Figure 2: Difference average scalograms of CWTs: (a) CP3 channel and (b) C3 channel (subject AA)

The local extreme $\left(a_i^k, b_j^k\right)$ of the two-sample t-statistic $t^k(a,b)$ are found and the $W^{kn}(a,b)$ is the extracted feature representing the $\overline{CWT_s^k}\left(a_i^k, b_j^k\right)$, that is all the values of the $\overline{CWT_s^k}(a,b)$ are 0, except those for which there were found extreme**.** At this step, we take into account a threshold equal to $t_{crit} = 1.6002$ [12], for 95% confidence interval. These points represent the local maxima difference between the two groups in the time-scale. In figure 3 the features extraction for top movement of the cursor for CP3 and C3 channel is shown.



(a)



(b)

Figure 3: Feature extraction (a) CP3 channel and (b) C3 channel (subject AA)

Conclusions

Choosing the right parameters of the wavelet is accomplished by means of computing first the Hurst coefficient and the fractal dimension of the signals. This led to a better possibility of feature extraction due to the lower correlation coefficients between the signals.

The combined two methods proposed for feature extraction, the continuous wavelet transform using complex Morlet wavelet and the Students' t-test, applied to EEG signals, pointed out the possibility of quantification of mu rhythms that best discriminate between the two opposite groups: one for mental imagination of moving the cursor to the top and the other to the bottom of one edge of a screen of a monitor.

References

[1] WALPAW J., MCFARLAND D., VAUGHAN T. (2000): Brain-Computer Inteface Research at the Wadsworth Center, IEEE Transactions on Rehabilitation Engineering, Vol. 8, no. 2, June, 222-226

ISSN: 1727-1983

[2] HURST H., (1951): Long-Term Storage Capacity of Reservoirs, Trans. Am. Soc. Civil Eng., 116, 770-799

[3] FADILI M.J., BULLMORE E.T., (2002): Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors, NeuroImage 15, 217-232

[4] BOSTANOV V., (2004) BCI Competition 2003-Data Sets Ib and IIb: Feature Extraction from Event Related Potentials with the Continuous Wavelet Transform and the t-Value Scalogram, IEEE Transactions on Biomedical Engineering, vol. 51, no. 6, June, 1057-1061

[5] BLANCHARD G., BLANKERTZ B, (2004): BCI Competition 2003: Data Set IIa – Spatial Patterns of Self-Controlled Brain Rhythm Modulations, IEEE Transactions on Biomedical Engineering, vol. 51, no. 6, June, 1062-1067

[6] DELOME A., MAKING S., (2003): EEG Changes Accompanying Learned Regulation of 12-Hz EEG Activity, IEEE Transactions on Rehabilitation Engineering, 11(2), 133-136

[7] WADSWORTH BCI Dataset NIPS*2001 BCI Classification Contest Documentation, http://ida.first.fhg.de/projects/bci/competition/albany_desc/albany_desc_i.html

[8] NINNESS B., (1998): Estimation of 1/f noise, IEEE Trans. on Information Theory, 44, 32-46

[9] LUX T., MARCHESI M., (1999): Scaling and criticality in a stochastic multi-agent model of a financial market, Nature, 397, 498-500

[10] ALDROUBI A., UNSER M., (1996) Wavelets in Medicine and Biology, CRC Press

[11] VALENS C., A Really Friendly Guide to Wavelets, http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html#eq1

[12] Student t-test, http://techniques.geog.ox.ac.uk/mod_2/week_5/lecture-5.htm

[13] WALPAW J., MCFARLAND D., NEAT G., FORNERIS CH., (1991): An EEG-Based Brain-Computer Interface for Cursor Control, Electroencephalography and Clinical Neurophysiology, 78, 252-259

ISSN: 1727-1983

ISSN: 1727-1983 © 2005 IFMBE