

Feature Extraction on Brain Computer Interfaces using Discrete Dyadic Wavelet Transform: Preliminary Results

I Gareis¹, G Gentiletti¹, R Acevedo^{1,2}, L Rufiner^{1,2}

¹ Laboratorio de Ingeniería en Rehabilitación e Investigaciones Neuromusculares y Sensoriales - Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina.

² Centro de I+D en Señales, Sistemas e Inteligencia Computacional - Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional de Litoral, Argentina.

E-mail: racevedo@bioingenieria.edu.ar

Abstract. The purpose of this work is to evaluate different feature extraction alternatives to detect the event related evoked potential signal on brain computer interfaces, trying to minimize the time employed and the classification error, in terms of sensibility and specificity of the method, looking for alternatives to coherent averaging. In this context the results obtained performing the feature extraction using discrete dyadic wavelet transform using different mother wavelets are presented. For the classification a single layer perceptron was used. The results obtained with and without the wavelet decomposition were compared; showing an improvement on the classification rate, the specificity and the sensibility for the feature vectors obtained using some mother wavelets.

1. Introduction

The technique most commonly employed to improve the Signal to Noise Ratio (SNR), and estimate the Evoked Potential signal (EP), is coherent averaging. It consists on applying successive stimuli and average the electrical activity registered synchronized with the time instant in which the stimulus is applied.

Nevertheless coherent averaging is of massive use, it presents some issues that make desirable the finding of alternatives to this technique. One of these issues is the time required to achieve an accurate estimation of the EP, another one is that it assumes that the EP remains stationary from trial to trial and that the noise is white with zero mean, which is not completely true in general. As a consequence the result obtained from the coherent averaging is a bad estimation of the EP signal [1].

On the other hand there are situations in which it is not necessary to estimate the signal but rather to detect if it is present or not. This is the case of Brain Computer Interfaces (BCI). There are many BCI paradigms that can be initially classified in two groups: a) Invasive: in this case the signals are provided by intracranial or intracortical registers and b) Non invasive: using surface registers, from which the most common are the ones based on electroencephalography (EEG). To this last group belongs the paradigm based on Event Related Potentials (ERP) [2].

ERP are potentials of latencies larger than 100 ms, and their manifestation depends of psychological and cognitive processes. The most studied ERP is the P300 signal. When visual, auditory or somatosensory infrequent or particularly meaningful stimuli are mixed with frequent stimuli, the first evoke a potential over the parietal cortex with a positive peak close to the 300 ms that can be

registered on the EEG. To estimate the ERP signal it is necessary to improve the initial SNR by means coherent averaging.

Donchin *et al* [3] were the first to use the p300 on a BCI with visual stimuli organized by rows and columns over a character matrix that allowed its spelling. The most outstanding advantage of this technique is that it does not require the user to train, and on the same time it results natural on target selection tasks *e.g.* word spelling, direction selection, menu selection, among others. Many works suggest that P300 based BCIs can achieve a data communication rate of 25 bits/min working online [4].

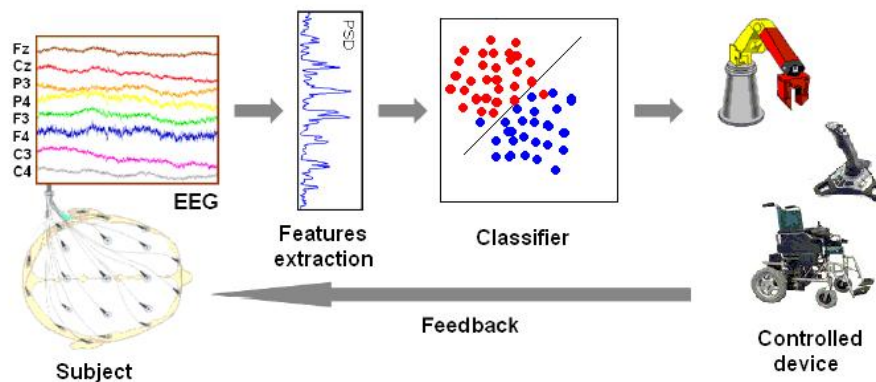


Figure 1. General architecture of a BCI for device control.

In Figure 1 is shown the general architecture of a BCI proposed by Millán *et al* [5], where its functional blocks are described, in which a subject controls a device, *e.g.* a wheelchair, a robotic arm or a joystick. The user can also check the device state through feedback that allows him to acknowledge the result of his efforts to control it.

Recently, with the advance in algorithms on machine learning and technologies of digital processing, a part of the investigation on BCI is the exploration of feature extraction techniques and signal classification.

On a BCI like the one described the following blocks can be clearly distinguished:

- a) Instrumentation block: is the acquisition and conditioning system for the electroencephalogram.
- b) Feature extraction block: its objective is to generate a suitable representation of the original signal (pattern in the context of Pattern Recognition) that allows improving the performance of the classifier.
- c) Classifier block: its function is to identify to which class does a pattern belongs to.

In this particular problem the last two blocks are a pattern recognition system where the classification problem to solve has two possible classes: registers with ERP and registers without ERP. The objective of this work is to evaluate different feature extraction alternatives to detect the ERP signal in BCI minimizing the time employed, and maximizing as well the accuracy, sensitivity and specificity of the method. We present here some preliminary results of applying a signal decomposition technique based on dictionaries of time-frequency atoms, which are generated using Wavelet multiresolution decomposition.

2. Methodology

2.1. ERP registers

In this work there were used registers obtained with amplifiers Grass® model 8-18-36, the acquisition parameters of the EEG signals are displayed on table I. The BCI2000 software developed in the Wadsworth Center (Albany, New York) [6] was used.

Table 1. EEG acquisition parameters.

Parameter	Description
Channels	Fz, Cz, Pz, Oz, C3, C4, M1 (reference) y M2 (ground)
Sample Frequency	1024 Hz

To minimize the amount of information presented to the classifier successive decimations were performed over the temporal patterns, with the objective of setting the amount by which the number of samples could be reduced without reducing the performance of the classifier. Care was taken to avoid the aliasing by applying a Chebyshev filter. The decimations were made by multipliers of two for convenience.

For the stimulation, the Donchin speller [3] was used, although the alphabet characters were replaced by icons that control the movement of a wheelchair [7].

2.2. Feature Extraction

There are many alternatives for the feature extraction block, some more suitable than others depending on the signal and the features that are intended to be emphasized. In this work we propose to use approximation methods of approximation based on dictionaries of discrete signals like the used in [8],[9]. In these methods the signal of interest is considered as an element of a signal space that can be represented in terms of a dictionary or base that remark meaningful features. In this context the signal $\mathbf{x} \in \mathbb{R}^N$ can be expressed as a function of an appropriate orthogonal base using the equation $\mathbf{a} = \boldsymbol{\phi} \cdot \mathbf{x}$ where $\mathbf{a} \in \mathbb{R}^N$ is the vector expressed in the new base and $\boldsymbol{\phi} \in \mathbb{R}^{N \times N}$ is a matrix which columns are the elements of the new base, also called atoms [8],[9]. In this work the vector \mathbf{a} was built by means of a multiresolution decomposition produced by the Discrete Dyadic Wavelet Transform (DDWT) using mother wavelets of different families which were selected by their "visual" resemblance with the P300 peak. In this process, the signal is separated in a portion of high frequencies (called detail) and a portion of low frequencies (called approximation), each with half of samples of the original signal, which represent the first level of decomposition. The decomposition in each posterior level is made only over the approximation of the previous one, obtaining another detail and approximation portions, with half the samples of the previous approximation.

Previous to the application of the DDWT the registers (signals) were filtered using a low pass Chebyshev digital filter of order eight with a cutoff frequency that depended on the subsequent subsampling, after this the registers were subsampled. After the DDWT different levels of details were selected with the approximation, with these coefficients there were formed vectors, one corresponding to each trial. These vectors formed the groups of patterns to present to the classifier. The coefficient groups were selected taking into account that the ERPs are formed by low frequency components, therefore the approximations were always kept and the amount of detail levels used were varied, always keeping the details of the bigger scales and adding detail coefficients of the smaller scales to be able to define which of the detail scales have useful information for the classification process.

2.3. Classifier

Assuming that the problem classes are linearly separable a single layer perceptron was used.

2.4. Performance

The indices chosen to evaluate the performance of classification were the accuracy, sensitivity and specificity, defined as follows:

$$\text{Accuracy} = \frac{\text{Correctly classified patterns}}{\text{Total of patterns}} \quad \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

where TP are true positives values, TN are true negatives values, FP are false positives values and FN are false negatives values.

3. Results

3.1. Registers

In Figure 2 are shown the 300 trials averages ERP obtained in simultaneous registers on positions F_z , C_z , P_z , O_z , C_3 y C_4 in their respective positions. It can be seen that the morphology of ERP depends on the position of the electrodes on which the signal is acquired, and that in some positions it manifests better than in others. From the six were selected two channels on which the P300 peak is manifested with a different polarity, this were channels C_z and O_z . On them it was evaluated how does the amount of trials to average influence the time frequency analysis.

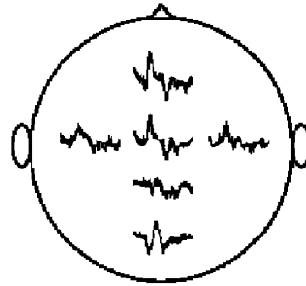


Figure 2. ERP waveform in different positions of the scalp.

Figure 3 shows the scalograms that result from applying Continuous Wavelet Transform (CWT), using Biorthogonal 3.9 as mother wavelet, in averages of 10, 5 and 1 trials to patterns with and without ERP from channels C_z and O_z . It can be seen that the scalograms show differences between the patterns with and without ERP, even though it is clearer on channel O_z . This situation is maintained noticeably for averages of up to three trials (not shown in this work) for channel O_z . Due to these we worked with data from channel O_z .

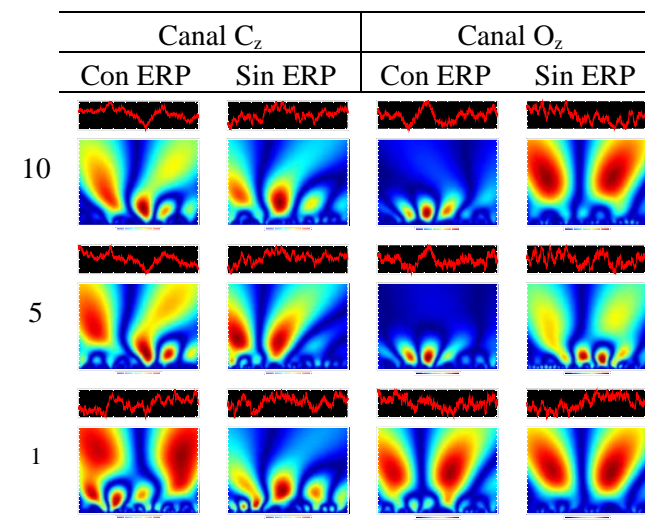


Figure 3. Scalograms of C_z and O_z channels with different amount of averaged trials

In Figure 4 the averages obtained from registers corresponding to channel O_z are shown; on blue line the average of tree hundred trials with ERP while on red line is displayed the average of tree hundred trials without ERP.

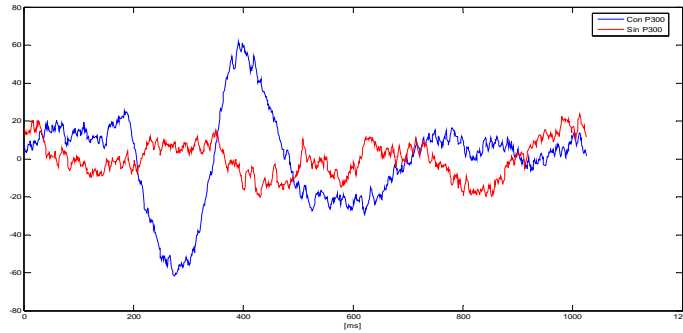


Figure 4. Averages with and without ERP on channel O_z .

3.2. Subsample

Table II summarizes the performance presenting single trial temporal decimated registers to the classifier. It can be seen that systems performance was maximum when decimating by four. Because of that decimation by four was set to be used previously to the application of the DDWT.

Table II. Systems performance for different decimation levels

Decimation	Accuracy	Sensitivity	Specificity
No decimation	63 %	0,7	0,56
x 2	63 %	0,72	0,54
x 4	64,75 %	0,755	0,54
x 8	61,75 %	0,69	0,54
x 16	62 %	0,7	0,53
x 32	62 %	0,69	0,55

3.3. ERP detection

Considering the results obtained in the time-scale analysis previously described, and the performance of the system for the decimated temporal patterns, sets of temporal patterns that consisting of single trials corresponding to the O_z channel were generated, all decimated four times. Resultant from the multiresolution decomposition of each one of the temporary decimated patterns, there were generated patterns consisting the wavelet coefficients corresponding to the approximation of level six and the details of levels one to six. Approximation of level six was always conserved, because by frequency considerations it was assumed that the in that level of scale, the patterns contained information on the presence or not of the ERP. Considering each detail and approximation has half the samples of the approximation in the previous level, starting with a signal of 256 samples, is arrived, the sixth approach has only four samples. In table III the amount of samples of the approximation in the sixth level and of each one of the details in different levels is detailed for the decimated patterns of 256 samples.

Table III. Amount of samples for different decomposition levels

Decomposition depth		Number of samples
Level 1	Detail	128
Level 2	Detail	64
Level 3	Detail	32
Level 4	Detail	16
Level 5	Detail	8
Level 6	Detail	4
	Approximation	4

From the decimated single trial patterns the decomposition was made with different wavelet families, obtaining in this form the wavelet coefficients corresponding to the approximation in the sixth level and the details of levels one to six. Then the coefficients corresponding to the sixth level approximation and to the details of different levels were taken, to form with these the patterns that were presented to the classifier.

The table IV presents the performance of the single layer perceptron in the detection of patterns corresponding to decompositions of the patterns with some representative mother wavelets of, taking in each case different amount of details.

Table IV. System performance for different wavelets and details coefficients combinations

Details	Wavelet: Biortogonal 3.9			Wavelet: Daubechies 9		
	Accuracy [%]	Sensitivity	Specificity	Accuracy [%]	Sensitivity	Specificity
1-6	61.00	0.68	0.54	63.25	0.71	0.55
2-6	61.50	0.71	0.52	63.75	0.72	0.55
3-6	59.50	0.69	0.5	65.25	0.70	0.60
4-6	63.75	0.75	0.53	66.25	0.72	0.60
5-6	65.00	0.77	0.53	69.25	0.74	0.64

Details	Wavelet: Coiflets 4			Wavelet: Symlets 8		
	Accuracy [%]	Sensitivity	Specificity	Accuracy [%]	Sensitivity	Specificity
1-6	61.00	0.70	0.52	63.00	0.73	0.53
2-6	64.25	0.75	0.54	64.5	0.73	0.56
3-6	66.75	0.75	0.58	62.50	0.72	0.53
4-6	64.00	0.76	0.52	65.75	0.76	0.55
5-6	64.75	0.75	0.55	60.50	0.72	0.49

4. Conclusions

From the analysis of results it can be seen that the best performance is superior when using Daubechies 9 as mother wavelet for the decomposition, this can be seen in general for every mother wavelet belonging to this family that were used. It can also be noticed that the system performance is lower when performing the decomposition with some wavelet families than using a temporal base. It is also important to highlight that the sensibility was in general bigger than the specificity, result that in case of being consistent must be taken into account on the brain computer interface design with the objective of improving its efficiency.

A similar analysis was also made using averages of 2, 3, 4, 5, 6, 7, 8, 9 and 10 trials reaching accuracy values of about 75% with averages of ten trials applying DDWT and almost of 70% presenting to the SLP the ten trial averages of temporal registers.

In spite of being an initial step in this line of work encouraging results have been achieved, since using a simple and probably not optimal classifier and using registers of only one channel with averages of two trials acceptable classification rates were achieved.

However, to achieve a practically exploitable development the performance of the system should be improved; to achieve this goal some possible directions to consider are the inclusion of information obtained from other channels and the experimentation with other classifiers. Another possibility consists in designing a method to estimate an optimal discriminative dictionary.

Acknowledgements

This work was supported by the Universidad Nacional de Entre Ríos (with PID 6101 and 6106), the Universidad Nacional de Litoral (with CAI+D 012-72), STIC-AMSUD program (09STIC01), the Agencia Nacional de Promoción Científica y Tecnológica (with PAE 37122, PAE-PICT-2007-00052) and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) from Argentina.

References

- [1] R. Acevedo, C. Bonell, G. Gentiletti, J. Aldonate, C. Mercuri, S. Escobar and J. Biurrun 2006. Potenciales evocados auditivos de tronco cerebral: Aspectos tecnológicos del registro *Revista Argentina de Bioingeniería* (SABI), Vol. 12 No. 1, pp. 10-17.
- [2] G. Gentiletti, C. Tabernig and R. Acevedo 2007. Interfaz cerebro-computadora: Estado del arte y desarrollo en Argentina *Revista Argentina de Bioingeniería* (SABI), Vol. 13 No. 1, pp. 13-19, ISSN: 0329-5257
- [3] L. A. Farrel and E. Donchin 1988. Taking off the top of your head: toward a methal prothesis utilizing event related brain potentials *Electroencephalography and Clinical Neurophysiology*, vol. 70, pp. 510-523.
- [4] R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan 2002. Brain-computer interfaces for communication and control *Clinical Neurophysiology*, vol. 113, pp.767-791.
- [5] J. del R. Millán, P. W. Ferrez, F. Galán and E. Lew 2007. Non-invasive brain-machine interaction *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, pp. 1-13.
- [6] <http://www.bci2000.org/BCI2000/Home.html>
- [7] G.G. Gentiletti Faenze, J.G. Gebhart, M. Richard, E. García Cossio, R.C. Acevedo and V. Medina Bañuelos. Robotics Simulation Platform for BCI Application: A Wheelchair Driving Example Using P300 Paradigm *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course*, Graz (Austria), 2008
- [8] Chen, Donoho and Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM REVIEW*, Vol. 43, No. 1, pp. 129-159, 2001.
- [9] H.L. Rufiner 2009, "Análisis y modelado digital de la voz: técnicas recientes y aplicaciones", Ed. UNL, Colección Ciencia y Técnica.