Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009

Recognition and Classification of P300s in EEG Signals by Means of Feature Extraction Using Wavelet Decomposition

S. Costagliola, B. Dal Seno and M. Matteucci

***Abstract*—In the last twenty years the understanding of the brain function and the advent of powerful low-cost computer equipment allowed the birth and the development of the BCI (Brain-Computer Interface), a device that interprets brain activity to issue commands. P300 is a positive peak at about 300 ms from a stimulus, and has been used as a base for a BCI in many studies.**

The aim of this research consists in recognizing and classifying P300 signals by using wavelet transforms. This study analyzes both the kind of wavelets and which coefficients are more suited for a 100% correct decisions using as few repetitions of stimuli as possible. The classifier performs a quadratic discriminant analysis. The method is tested on the "BCI Competition 2003" data set IIb with excellent results.

I. Introduction

To this day many people still suffer from a wide variety of neuromuscular disorders, which include amyotrophic lateral sclerosis (i.e., ALS or Lou Gehrig's disease), multiple sclerosis, cerebral palsy and spinal cord injury. These pathologies can interrupt or otherwise impair the pathways that control motoneurons in the spinal cord and brainstem, disrupting the channels through which the brain normally communicates with and controls its external environment.

In the absence of means for repairing the damaged nervous system, three options exist for restoring function. The first is to augment the capabilities of the remaining pathways; the second is to detour around points of damage and the third is to provide the brain with wholly new channels for communication and control, the so called BCI (brain-computer interface).

This third way, unlike the other two, has remained largely unexplored until the last twenty years, during which research suggested that electroencephalographic (EEG) activity can provide the basis for such new channels.

EEG activity occurs continuously in both humans and animals; however, if EEG activity is recorded in relation to a specific stimulus, one of its components is referred to as an event-related potential (ERP). ERPs are superpositions of the electrical activities of millions of neurons that reflect the response of the brain to internal or external stimuli. In a communication context this latter can represent our information signal called target stimulus. The response to the target stimulus elicits the P300 that manifests itself as a positive voltage approximately 300 milliseconds (ms) after the stimulus is presented. This is one of the most robust features of the ERP response, which is commonly investigated in behavioral neuroscience.

S. Costagliola, M. Matteucci and B. Dal Seno are with Politecnico di Milano, Dipartimento di Elettronica e Informazione, IIT unit, Via Ponzio 34/5, 20133 Milano, Italy (e-mail: costagliola@elet.polimi.it).

II. P300 and the "Oddball Paradigm"

The P300 has a number of desirable qualities that aid in the implementation of BCI systems. First, the waveform is consistently detectable and is elicited in response to precise stimuli. The P300 waveform can also be evoked in nearly all subjects with little variation in measurement techniques, which may help simplify interface designs and permit greater usability.

P300 is most frequently related to the framework of what has come to be called the "oddball paradigm" [1]. In this paradigm the subject is presented with a sequence of events that can be classified into two categories. In general, events in one of the two categories are rarely presented. Furthermore, the subject is assigned a task that cannot be performed without categorizing the events. Under these circumstances, events in the rare category elicits an ERP characterized by a P300 component; the less probable the eliciting event, the larger the P300.

Farwell and Donchin [2] described a BCI that exploited these properties of the oddball paradigm to allow a user to communicate a sequence of letters to a computer. An oddball paradigm was created by successively, and randomly, intensifying either a row or a column of a 6 by 6 matrix of characters that was displayed continually to the subject (figure 1).

In each "trial" the subject is "communicating" a character by focusing attention on the cell containing the character. Hence, the total sequence of events is divided into two categories. One category, which constitutes 16.7% of the intensifications (one in six), includes the cell whose content are at the focus of attention. The remaining intensifications are of rows and columns that do not include the relevant cell. If this is indeed an oddball paradigm, then the events containing the relevant cell (being the rare events in an oddball paradigm) should be the only events that elicit a P300.

The communication task reduces, thus, to the detection of which row, and which column, are those eliciting a P300 on a given trial. The letter the subject is trying to communicate is at the intersection of the row and the column that elicited a P300.

The speed of the highlighting determines the number of

978-1-4244-3553-1/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.

characters processed per minute but the real speed at which an interface is able to operate depends on how detectable the signal is despite "noise". One negative characteristic of the P300 is that the low amplitude of the waveform requires averaging of multiple recordings to isolate the signal. This and other post-recording processing steps determine the overall speed of an interface [3]. Results from studies using this setup show that normal subjects could achieve a 95% success rate at 3.4-4.3 chars/min. It remains to be shown whether such systems provide similar results in patients suffering from "locked-in" syndrome, the main target population for such brain-driven devices.

Fig. 1. The stimulus matrix monitored by the subject. Every 175 ms (5.7 Hz) one of the rows, or one of the columns of the matrix was intensified.

III. Wavelet-based feature extraction

Since ERPs are non-stationary signals, time-domain filters as well as frequency-domain filters cannot deal with single trial variabilities. Among various filtering techniques described in the literature, the discrete wavelet transform (DWT) is potentially one of the most powerful techniques. This is because of its ability to adjust to signal components, as well as the speed of its computation. It provides a time-frequency decomposition that is shown to be very suitable for ERP analysis, due to its optimal resolution both in the time and in the frequency domain [4] [5].

The wavelet transform (WT) is defined as the scalar product between the signal $x(t)$ and the wavelet functions $\psi_{s,\tau}(t)$:

$$W(s,\tau) = \langle x(t), \psi_{s,\tau}(t) \rangle \tag{1}$$

where $\psi_{s,\tau}(t)$ are dilated (contracted) and shifted versions of a unique wavelet function (mother wavelet) $\psi(t)$,

$$\psi_{s,\tau}(t) = s^{-1/2}\psi\left(\frac{t-\tau}{s}\right) \tag{2}$$

(s, τ are the scale and translation parameters, respectively). The WT gives a decomposition of $x(t)$ in different scales, tending to be maximum at those scales and time locations where the wavelet best resembles $x(t)$. Moreover, eq. (1) can be inverted, thus giving the reconstruction of $x(t)$. The WT maps a signal of one independent variable t onto a function of two independent variables s, τ. This procedure is redundant and not efficient for algorithmic implementations. In consequence, it is more practical to define the WT only at discrete scales s and discrete times τ by choosing the set of parameters $\{s_j = 2^j;\ \ \tau_{j,k} = 2^j k\}$, with integers j, k. Contracted versions of the wavelet function match the high-frequency components of the original signal and on the other hand, the dilated versions match the low frequency oscillations. Then, by correlating the original signal with wavelet functions of different sizes we can obtain its details at different scales. These correlations with the different wavelet functions can be arranged in a hierarchical scheme called multiresolution decomposition using multirate filterbanks [6]. This outlook shows the DWT as a constant Q filterbank with octave spacing between the centers of the filters. Each subband contains about half the samples of the neighboring higher frequency subband. In the pyramidal algorithm the signal is analyzed at different frequency bands with different resolution by decomposing the signal into a coarse approximation and detail information. The coarse approximation is then further decomposed using the same wavelet decomposition step. Because of the downsampling the number of resulting wavelet coefficients is approximately the same as the number of input points. However, in the wavelet domain the most of energy of the P300 is localized in few coefficients. The choice of these coefficients joined with the power of the classification method are the two factors that determine the success of the classification.

IV. Classification

From a mathematical point of view, the P300 detection problem can be addressed by considering two different classes $k = \{1,2\}$. The first one includes all the signals where the related raw/column is intensified and it represents 1/6 of all cases whereas the second one includes the remaining 5/6 where the ERP is not present.

Following the Bayes theorem we can compute the posterior probabilities of one class given the specific realization x:

$$p(K = k \mid X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{2} f_l(x)\pi_l} \tag{3}$$

where $f_k(x)$ are the class-conditional densities of data X

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.

in class $K = k$ and π_k are prior probabilities of the two classes (1/6 and 5/6 respectively). The class that maximize the posterior probability (MAP) can be estimated by:

$$\begin{aligned}\hat{K}(x) &= \arg\max_k p(K = k \mid X = x) \\ &= \arg\max_k f_k(x)\pi_k\end{aligned} \tag{4}$$

If the distribution of the wavelet coefficients is Gaussian we can write class-conditional density as a multivariate Gaussian:

$$f_k(x) = \frac{1}{(2\pi)^{N/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)} \tag{5}$$

where N is the dimensionality of x and μ_k and $\mathbf{\Sigma}_k$ are the mean vectors and the covariance matrices of the two classes respectively. These are all that we need to train the classifier. We can estimate mean vectors and covariance matrices directly from the training set:

$$\begin{cases} \hat{\mu}_k = \sum_{K_i = k} \frac{x_i}{N_k} \\ \hat{\mathbf{\Sigma}}_k = \sum_{K_i = k} \frac{(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T}{N-1} \end{cases} \tag{6}$$

Under these assumptions decision boundaries are quadratic equation in x hence the name "Quadratic Discriminant Analysis" (QDA). The letter the subject is trying to communicate is estimated to be at the intersection of the row and the column that are finally chosen (independently) maximizing the simple functionals (see appendix A for a derivation of the formulas):

$$\begin{cases} row = \arg\max_{x_{r\{1\to6\}}} \left\{ x_r^T \left[\hat{\mathbf{\Sigma}}_{r1}^{-1}\left(\mu_{r1} - \frac{x_r}{2}\right) - \hat{\mathbf{\Sigma}}_{r2}^{-1}\left(\mu_{r2} - \frac{x_r}{2}\right) \right] \right\} \\ col = \arg\max_{x_{c\{1\to6\}}} \left\{ x_c^T \left[\hat{\mathbf{\Sigma}}_{c1}^{-1}\left(\mu_{c1} - \frac{x_c}{2}\right) - \hat{\mathbf{\Sigma}}_{c2}^{-1}\left(\mu_{c2} - \frac{x_c}{2}\right) \right] \right\} \end{cases} \tag{7}$$

V. Data Collection

The dataset used for testing the algorithm represents a complete record of P300 potentials acquired by the group of J. Walpow at the Wadsworth Center with BCI2000[1] for the "BCI Competition 2003" [7].

The signals (digitized at 240Hz) are recorded from one subject in three sessions. Each session consists of a number of runs. In each run, the subject focused attention on a series of characters (see table 1 for the target word for each run in session 10, 11 and 12). For each character, user display is as follows: the matrix is displayed for a 2.5 s period and during this time each character has the same intensity (i.e., the matrix is blank). Subsequently, each row and column in the matrix is randomly intensified for 100 ms (i.e., resulting in 12 different stimuli – 6 rows and 6 columns). After intensification of a row/column, the matrix is blank for 75 ms. Row/column intensifications are randomized in blocks of 12. Sets of 12 intensifications are repeated 15 times for each character (i.e., any specific row/column are intensified 15 times and thus there were 180 total intensifications for each character). Each sequence of 15 sets of intensifications are followed by a 2.5 s period, and during this time the matrix is blank. This period informed the user that this character is completed and to focus on the next character in the word that is displayed on the top of the screen (the current character is shown in parentheses). EEG data are recorded on a 64 channel system. The electrodes configuration is shown in figure 2.

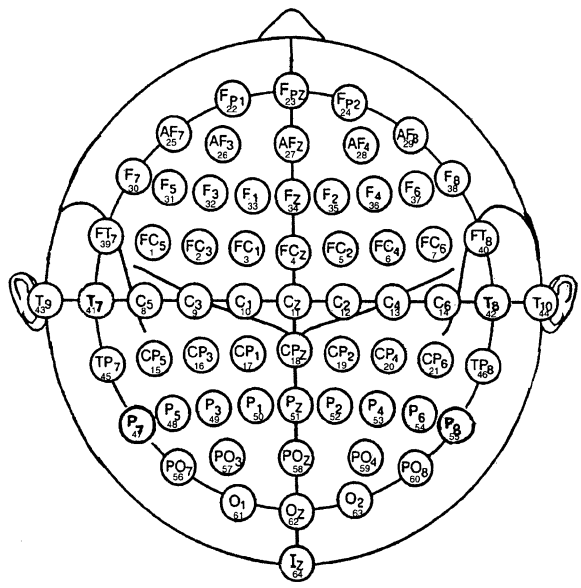

Fig. 2. Electrode designations (Sharbrough, 1991) and channel assignment numbers as used in the experiment.

TABLE I
WORDS THAT THE SUBJECT WAS FOCUSING

Run	Session 10	Session 11	Session 12
1	CAT	HAT	FOOD
2	DOG	HAT	MOOT
3	FISH	GLOVE	HAM
4	WATER	SHOES	PIE
5	BOWL	FISH	CAKE
6		RAT	TUNA
7			ZYGOT
8			4567

The goal in this competition is to use the labeled data in session 10 and 11 to train a classifier and then to predict the words in session 12.

[1] BCI2000 is a flexible Brain-Computer Interface research and development platform. It supports a variety of brain signals, signal processing methods and user applications. It is available free of charge for research purposes (http://www.bci2000.org).

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.

The submission with the highest percentage of correct letters won the competition. When the 100% of correct classifications is reached the final goal is to reduce the number of repetitions from 15 to the minimum possible. The algorithm presented is tested using the same rules.

VI. Data Analysis and Results

One of the most important factors that determine the success of the classification is how well the distinguishing pattern is represented by the features extracted from the training set; therefore a good choice of the features is essential. In this study wavelet coefficients are used as features; hence the choice of the mother wavelet is the first issue to deal with. In literature different kinds of wavelets are used. In [8] quadratic bi-orthogonal B-Splines (bior2.2) are chosen due to their similarity with the event-related responses and due to their optimal time-frequency resolution. In [9] Daubechies 4 (db4) wavelet is used because its shape is similar to the P300 waveform and because it possesses the exact reconstruction property. However, in our case the last characteristic is needless because we don't want to anti-transform and reconstruct the signal (i.e., for ERP denoising). Without this constraint a wider choice of wavelets is available.

We have therefore chosen to test many different types of wavelets, and select those that achieved the best results on the training set. In this way, the wavelet selection is done based on real data and it is therefore more likely to generalize well than a-priori choice that does not take into account the particular shape of the P300 of different subjects. An example of a mother wavelet for each considered family is shown in figure 3 (see [10] for details).

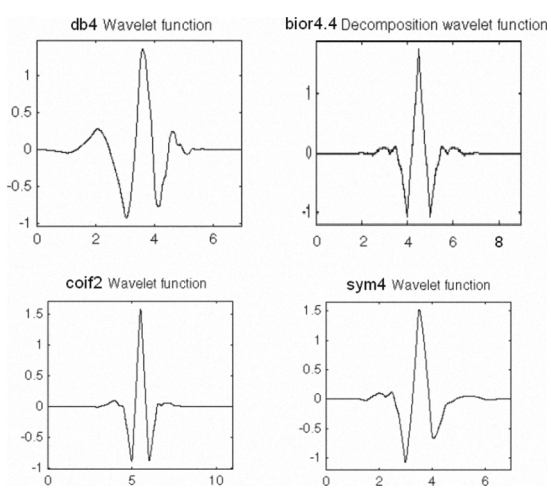

Fig. 3. Examples of mother wavelets.

Once the wavelet is chosen, 128 samples (corresponding to 533ms) from the start of the stimulus are extracted from the data coming from ten electrodes (the subset described by Kaper in [11] is used): Fz, Cz, Pz, Oz, C3, C4, P3, P4, Po7, and Po8.

A five-octave wavelet transform is performed yielding five sets of coefficients in the 60-120 Hz, 30-60 Hz (gamma), 15-30 Hz (beta), 8-15 Hz (alpha) and 4-7 Hz (theta) frequency ranges and the residues in the 0.5-4 Hz frequency band. However, the most pronounced features corresponding with the P300 wave are contained in the two lowest bands (about 20 coefficients depending on the wavelet). Therefore, the feature analysis is limited within a 20 by 10 matrix. In order to chose the dominant features correlating to the P300 wave, for each element of the matrix a t-test is performed, and the result is thresholded at the 70% of the total energy. The final number of features is hence reduced from 200 to about 20-30 (figure 4).

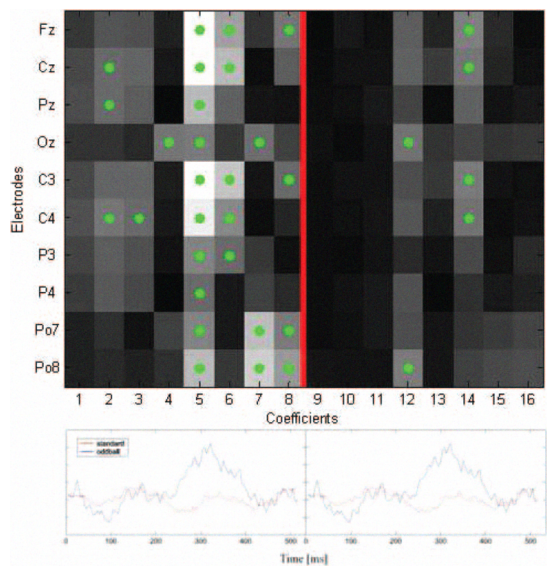

Fig. 4. Coefficients t-values. Green dots mark features above the threshold. The red line separate coefficients (1-8) in the lower band (0.5-4Hz) from the ones (9-16) in the higher band (4-7Hz). Time course of the P300 amplitude is represented below for comparison.

Before feature classification a statistical distribution test is done (an example of the two classes distributions is shown in figure 5). The wavelet coefficients distribution has an excess kurtosis of about 0.5 that depends on few outliers (spikes). Those outliers are discarded before the estimation of the statistical proprieties.

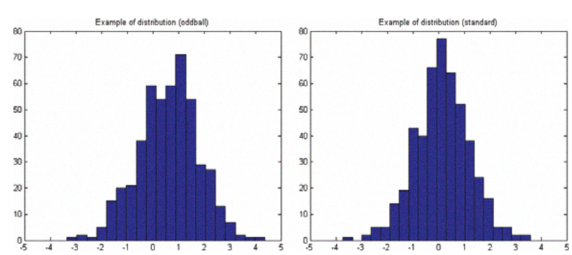

Fig. 5. Example of feature distribution (with and without P300)

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.

After this operation the test of Gaussianity gives a final excess kurtosis of about 0.1, which means that the coefficients distribution is quasi-Gaussian.

Once features are selected, those from the training set are used to parameterize the classifier by their mean vectors and covariance matrices (μ_k and $\mathbf{\Sigma}_k$). The analysis is performed by considering rows and columns separately thereby four vectors and four matrices are computed.

Both the mother wavelet and the threshold are chosen following up the classification performances obtained by applying equation (7) on a validation set. The validation set is composed of data coming from Session 10 Run 3,4 and Session 11 Run 3 for a total of 14 letters (about 30% of the whole training set). Results are shown in figure 6. Half of the tested wavelets reach the 100% of correct decision within few repetitions. In particularly db4, coif2 and sym4 reach a stable 100% within three repetitions, thereby these are selected for the P300 detection on the test set.

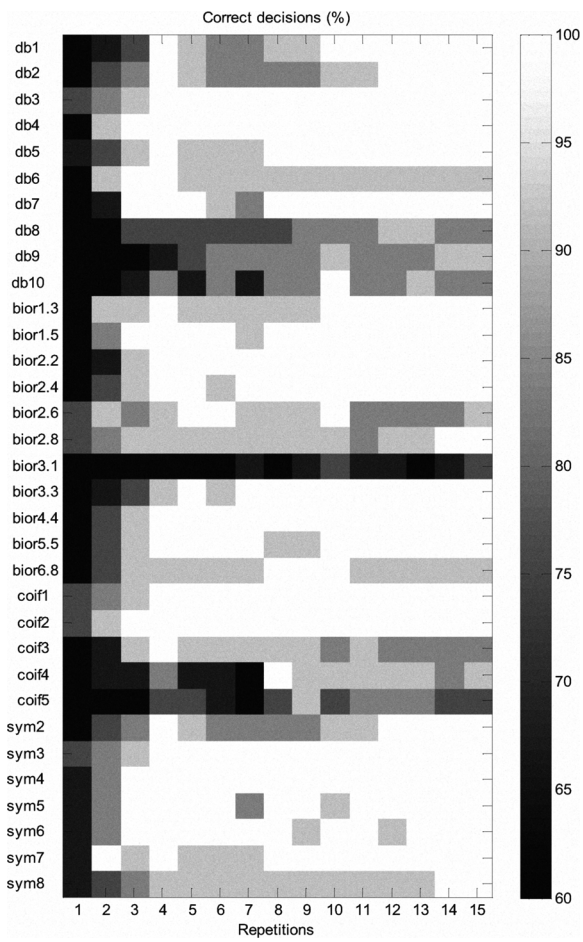

Fig. 6. Percentages of correct decisions on the validation set as a function of repetitions for each type of mother wavelet.

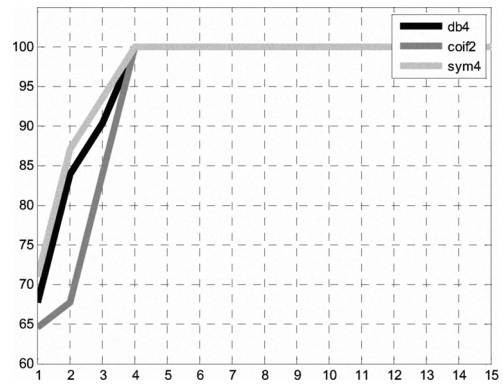

Fig. 7. Percentages of correct decisions on the test set as a function of repetitions for the selected mother wavelets.

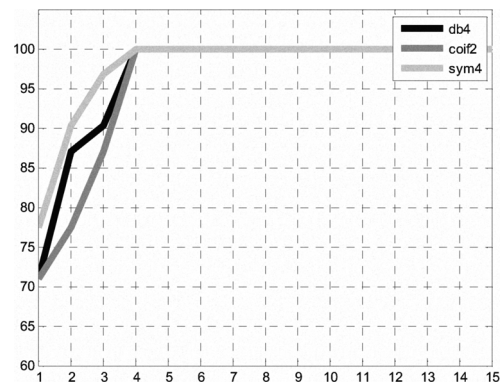

Fig. 8a. Percentages of correct row decisions on the test set as a function of repetitions for the selected mother wavelets.

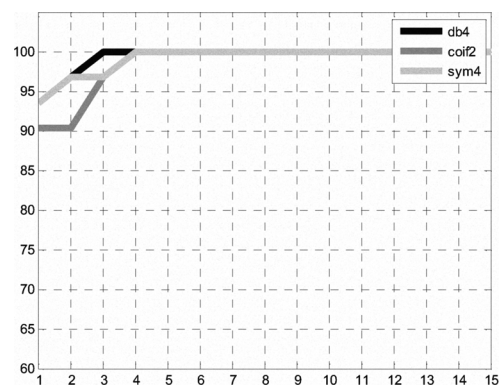

Fig. 8b. Percentages of correct column decisions on the test set as a function of repetitions for the selected mother wavelets.

Final results on the test set are shown in figure 7. The chosen wavelets allow to reach a stable 100% within four repetitions[2]. Row and column estimation performances are also shown separately in figures 8a and 8b. These two graphs highlight an interesting fact: in this data set the percentage of errors in column estimation is less than the

[2] Other wavelets give the same performance on the test set, though they have not been chosen because their performance on the validation set was lower. The 100% of correct decision with four repetitions is achieved using: db4, bior2.4, bior4.4, bior5.5, coif2, sym4 and sym6

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.

percentage of errors in row estimation. The same happens also in computing the probability of correct detection written in appendix (A-7). Results are shown in table 2. The probability of choosing the correct column on the first guess is about 13% more than the same probability on the row. Thereby, the idea of considering rows and columns separately turns out to be a remarkable choice.

VII. Discussion

By means of the wavelet transform, oscillatory transient features in the EEG signals are detected and the distinct functional components of P300 are extracted. The research among different types of wavelet points out that the most suited DWTs for the P300 detection are the following: db4, bior2.4, bior4.4, bior5.5, coif2, sym4, and sym6. With these wavelets a total of just 20-30 coefficients are extracted for representing the signals coming from all the electrodes in each intensification. With such few features 100% correct decisions is achieved on the "BCI Competition 2003" dataset even with only 4 repetitions of stimuli (the correct column is detected having even less repetitions).

Under the Gaussian assumption of the wavelet coefficients, the optimal classifier is derived. Its discriminant formula is simple and the computation is very fast, which makes it particularly suitable for online applications.

The method presented in this work achieve the same performance of the competition winners (actually, our results are slightly better, as we get 100% letters correct with fewer repetitions). Further experiments on different data sets are needed in order to compare the methods in a more robust way. We can affirm the presented method can be used as the basis for an online BCI or a new and better classification method.

Appendix-A

A derivation for equations (7) is presented in this appendix.

Let $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ be respectively the mean vectors and the covariance matrices of two possible classes which represent the observed vector $\boldsymbol{x}$. Under a Gaussian assumption the optimal classification involves the maximization of the posterior probabilities (MAP):

$$\begin{aligned}\hat{K}(x) &= \arg\max_k p(K=k \mid X=x) \\ &= \arg\max_k f_k(x)\pi_k \\ &= \arg\max_k \log(f_k(x)\pi_k) \\ &= \arg\max_k \begin{bmatrix} -\log\left((2\pi)^{N/2}|\Sigma_k|^{1/2}\right) + \log(\pi_k) + \\ -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) \end{bmatrix}\end{aligned} \tag{A-1}$$

where $f_k(x)$ are the class-conditional densities of X and π_k are prior probabilities of the classes.

Since just two classes are present, we can explicate the decision boundary:

$$\begin{aligned}&\log\frac{\pi_1}{\pi_2} - \frac{1}{2}\left(\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} + \mu_1^T\boldsymbol{\Sigma}_1^{-1}\mu_1 - \mu_2^T\boldsymbol{\Sigma}_2^{-1}\mu_2\right) + \\ &+ x^T\left(\boldsymbol{\Sigma}_1^{-1}\mu_1 - \boldsymbol{\Sigma}_2^{-1}\mu_2\right) - \frac{1}{2}x^T\left(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}\right)x = 0\end{aligned} \tag{A-2}$$

Without any other prior information we should decide that if the quantity on the left side is positive $\hat{K}=1$, otherwise $\hat{K}=2$. Conversely if we know that there must be (just) one vector $\boldsymbol{x}_j$ among M vectors $\boldsymbol{x}_i$ that belong to one class (i.e. the first one) we can estimate the index $\hat{j}$ maximizing over the data:

$$\begin{aligned}\hat{j} &= \arg\max_{x_i}\left\{\begin{array}{l} -\frac{1}{2}\left(\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} + \mu_1^T\boldsymbol{\Sigma}_1^{-1}\mu_1 - \mu_2^T\boldsymbol{\Sigma}_2^{-1}\mu_2 + x_i^T\left(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}\right)x_i\right) + \\ + x_i^T\left(\boldsymbol{\Sigma}_1^{-1}\mu_1 - \boldsymbol{\Sigma}_2^{-1}\mu_2\right) + \log\frac{\pi_1}{\pi_2} \end{array}\right\} \\ &= \arg\max_{x_i}\left\{x_i^T\left(\boldsymbol{\Sigma}_1^{-1}\mu_1 - \boldsymbol{\Sigma}_2^{-1}\mu_2\right) - \frac{1}{2}x_i^T\left(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}\right)x_i\right\}\end{aligned} \tag{A-3}$$

where the simplification is allowed because of the x-independence of the first terms. Therefore the functional that must be maximize is prior probabilities-independent and the classification is reduced to a ML (maximum likelihood) estimation. Finally, gathering terms, we obtain:

$$\hat{j} = \arg\max_{x_i}\left\{x_i^T\left[\boldsymbol{\Sigma}_1^{-1}\left(\mu_1 - \frac{x_i}{2}\right) - \boldsymbol{\Sigma}_2^{-1}\left(\mu_2 - \frac{x_i}{2}\right)\right]\right\} \tag{A-4}$$

We can reach the same result considering directly the joint probabilities composed by all the possible class-conditional densities of X:

$$p_i = p(K=1 \mid X=x_i)\prod_{l\neq i} p(K=2 \mid X=x_l) \tag{A-5}$$

The ML estimation become:

$$\begin{aligned}\hat{j} &= \arg\max_i p_i \\ &= \arg\max_i\left(f_1(x_i)\prod_{l\neq i} f_2(x_l)\right) \\ &= \arg\max_i\left[x_i^T\boldsymbol{\Sigma}_1^{-1}\left(\mu_1 - \frac{x_i}{2}\right) + \sum_{l\neq i} x_l^T\boldsymbol{\Sigma}_2^{-1}\left(\mu_2 - \frac{x_l}{2}\right)\right]\end{aligned} \tag{A-6}$$

that is equal to (A-4) because it differs just due to the common term $+\sum_{m=1}^{M} x_m^T\boldsymbol{\Sigma}_2^{-1}\left(\mu_2 - \frac{x_m}{2}\right)$.

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.

The probability of taking a correct decision $p_{\hat{j}=j}$ can be obtained from:

$$\begin{cases} Ap_{\hat{j}=j} = e^{-\frac{1}{2}\left[(x_j-\mu_1)^T \Sigma_1^{-1}(x_j-\mu_1)-(x_j-\mu_2)^T \Sigma_2^{-1}(x_j-\mu_2)\right]} \\ A = \sum_i Ap_{i=j} \end{cases} \quad \text{(A-7)}$$

where A is the scaling factor that normalize the probability.

ACKNOWLEDGMENTS

This work has been partially supported by the grant "Brain-Computer Interfaces in Everyday Applications" from Politecnico di Milano and Regione Lombardia.

REFERENCES

[1] M. Fabiani, G. Gratton, D. Karis and E. Donchin, "*Definition, identification, and reliability of the P300 component of the event-related brain potential*", Advances in Psychophysiology, P. K. Ackles, J. R. Jennings, and M. G. H. Coles, Eds. New York: JAI Press, 1987, vol. 2.

[2] L. A. Farwell and E. Donchin, "*Talking off the top of your head: A mental prosthesis utilizing event-related brain potentials*", Electroencephalogr. Clin. Neurophysiol., vol. 70, pp. 510–523, 1988.

[3] E. Donchin, K. M. Spencer, R. Wijesinghe, "*The Mental Prosthesis: Assessing the Speed of a P300-Based Brain–Computer Interface*", IEEE Transactions on Rehabilitation Engineering, 8(2) Jun 2000

[4] S. J. Schiff, A. Aldrouby, M. Unser, S. Sato, "*Fast wavelet transformation of EEG*", Electroencephalography and Clinical Neurophysiology, vol. 91, no. 6, pp. 442-455, December 1994

[5] R. Quian Quiroga, O. Sakowitz, E. Basar, M. Schuermann, "*Wavelet Transform in the analysis of the frequency composition of evoked potentials*". Brain Research Protocols, vol. 8, pp. 16-24, 2001/8.

[6] S. Mallat, "*A theory of multiresolution signal decomposition: the wavelet representation*", IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 No. 7, July 1989, pp. 674-693

[7] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer. "*The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials*". IEEE Transactions on Biomedical Engineering, 51(6):1044-1051, 2004

[8] R. Quian Quiroga and H. Garcia, "*Single-trial event-related potentials with wavelet denoising*", Clin. Neurophysiol. 114 (2003), pp. 376–390.

[9] S. A. Markazi, S. Qazi, L. S. Stergioulas, A. Ramchurn, D. Bunce, "*Wavelet Filtering of the P300 Component in Event-Related Potentials*". Engineering in Medicine and Biology Society, EMBS 2006. 28th Annual International Conference of the IEEE, 2006:1719–22.

[10] M. Misiti, Y. Misiti, G. Oppenheim, J. M. Poggi, "*Wavelet Toolbox™ User's Guide*", The MathWorks, Inc. 2008

[11] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "*Support vector machines for the P300 speller paradigm*". IEEE Transactions on Biomedical Engineering 51(6):1073–1076.

Authorized licensed use limited to: Politecnico di Milano. Downloaded on December 4, 2009 at 17:48 from IEEE Xplore. Restrictions apply.