Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics, June 12-15, Noordwijk, The Netherlands

Single-Trial EEG Classification of Movement Related Potential

Gabriel Pires, Urbano Nunes and Miguel Castelo-Branco

***Abstract*— A single trial electroencephalogram (EEG) classification system is proposed for left/right self-paced tapping discrimination. Features are extracted from *theta*, *mu* and *beta* rhythms and Readiness Potential (Bereitschaftspotential) that precede the voluntary movement. Feature extraction relies on regression fitting and wavelet decomposition. These two approaches are compared through two linear classification functions, a *Fisher Linear Discriminant* and a *Minimum-Squared-Error Linear Discriminant Function*. We show that discrete wavelet decomposition is an effective tool for both EEG frequency component separation and feature extraction, and therefore suitable for pre-movement left/right discrimination. The algorithms are applied to the data set <selfpaced2s> of the "BCI Competition 2001" with a classification accuracy of 96%.**

I. INTRODUCTION

Discrimination of user intentions mind reading from EEG sequences is a really challenging research field. For people suffering from progressive neuro-degeneration that causes loss of motor neurons leading to a complete paralysis such as amyotrophic lateral sclerosis (ALS), Brain Computer Interface (BCI) is a promising technique that can enhance the quality of living of these patients. Extensive research has been done in the classification of single-trial classification of EEG sequences recorded with non-invasive techniques. In [1] the current main BCI paradigms/experiments and associated research groups are presented. One of the experiments, called self-paced tapping, consists on the classification of pre-voluntary movement of left and right hand fingers. Notwithstanding the movement is actually performed, recorded data do not include the movement moment. Patients suffering from partial or complete paralysis still exhibit pre-voluntary movement potentials and are able to control their Slow Cortical Potentials (SCP) through motor imagery, suggesting their potential use for BCI [2]. In this paper we propose to discriminate the source of motor movements (left and right hand fingers), based on features extracted from Bereitschaftspotential (BP) and Event Related Desynchronization (ERD) in *mu* and *beta* rhythms of single trial EEG sequences. This experimental paradigm has already been extensively researched by some authors. Feature extraction, such as Autoregressive models and Common Spacial Subspace Decomposition, and classification techniques based on Neural Networks, Linear Discrimination and Kernel functions are discussed in [3], [4], [5] [6]. In [7], a new wavelet called SNAP was designed to match neural activity underlying the neuroelectric events. The coefficients resulting from this wavelet were used as features for classification, presenting better accuracy than with other known wavelets.

This paper makes a full frequency and time domain characterization of BP and ERD rhythms, and presents wavelets, specifically discrete wavelets, as an effective tool for feature extraction and component separation of neuroelectric waveforms. Notwithstanding the wide use of wavelets in several research areas, includig EEG analysis, they are not commonly used as a tool for feature extraction specifically in BCI. However, the non-stationarity properties of EEG patterns require feature extraction techniques capable of draw out temporal information. Wavelet based approaches have such characteristics and also provide frequency information. This approach provides a good classification accuracy when compared with other feature extraction methods and requires low computational speed. Two linear classifiers, a *Fisher Linear Discriminant* (FLD) and a *Minimum-Squared-Error Linear Discriminant Function* (MSE-LDF) confirm feature effectiveness.

Gabriel Pires is with the Institute for Systems and Robotics (ISR), University of Coimbra, Portugal gpires@isr.uc.pt

Urbano Nunes is with the Institute for Systems and Robotics (ISR), University of Coimbra, Portugal urbano@isr.uc.pt

Miguel Castelo-Branco is with the Biomedical Institute for Research in Light and Image (IBILI), University of Coimbra, Portugal mcbranco@ibili.uc.pt

II. METHODOLOGY

A. Experimental Paradigm and Data set

The EEG data set was made available by [4] for the NIPS*2001 (BCI) post-workshop competition. The experiment (<selfpaced2s>) [8] is as follows. A subject is seated in a normal chair with relaxed arms resting on the table and, in a self-chosen order, presses a keyboard key with fingers of either left or right hand (a complete description can be found in [4]). Each single trial measurement was recorded from 27 Ag/AgCl electrodes positioned according to the 10-20 international system. The EEG sequences correspond to data recorded from 1620 ms to 120 ms before the respective key press. For a sampling rate of 100 Hz, this corresponds to 151 data points. The data set is composed by a total of 516 trials, corresponding to 219 labeled *left* events, 194 labeled *right* events and 100 *test* events. The data set test is currently available with labels, but was not at the moment of the competition.

B. Bereitschaftspotential and ERD

The Bereitschaftspotential (readiness potential) is a negative shift of Slow Cortical Potential that precedes the voluntary movement. This is a DC negative shift that starts to be more pronounced about 500 ms before the onset of the movement [9]. For finger and hand movements, the BP is characterized by a contralateral dominance, and is more obvious at the lateral scalp positions C3 and C4 of the international 10-20 system, which corresponds respectively to the left and right primary motor cortex [10] [3].

The amplitude decrease and increase of cortical rhythms related to an external or internal event are known respectively, as event related desynchronization (ERD) and event related synchronization (ERS). The ERD oscillatory rhythms start over the contralateral and ipsilateral sensorimotor area shortly before the movement onset, contrarily to BP that appears 1 *s* before.

C. Feature extraction

The BP DC shift only becomes apparent after averaging a large number of trials. Fig. 1a) shows the average over all trials of potentials measured in $C1$, $C2$, $C3$ and $C4$ scalp positions. The contra-lateral dominance is more apparent in channels $C2$, $C3$ and $C4$ and therefore suitable for left/right discrimination. Simple functions of discrimination can be $y = C4 - C3$ or, using the three channels:

$$y = \frac{C4 + C2}{2} - C3 \tag{1}$$

The function (1) is used here like it was in [3], since it reveals slightly better results in left/right discrimination. A pre zero-mean normalization is applied in both cases. The function applied to the average over all trials and to a single trial is plotted in Fig. 1b). The single trial presents a large variance when compared with the average, which hides the potential shift. The power spectrum characterization reveals a contra-lateral dominance and, as expected, the signal information of the averaged trials is contained within the range DC-5Hz (Fig. 1c)). On the other hand, in the single trial, chosen at random, plotted in Fig. 1d), the DC-5Hz frequency component exists, but it is no longer the main component. The rhythms *theta* (4-8 Hz), *mu* (8-12 Hz) and *beta* (18-24 Hz) prevail and the contra-lateral dominance is not clear either in the overall average neither in the single trial. This foresees a difficult detection of the BP. To investigate the existence of ERD rhythms, the $C3$ and $C4$ averages were respectively band-pass filtered with 10 Hz and 20 Hz center frequencies (Fig. 1e)-g)). The frequency spectrum in *mu* and *beta* rhythms denotes a contra-lateral power reduction, which suggests a desynchronization. Time domain representations of *mu* and *beta* rhythms depicted in Fig. 1f)-h) show a *mu* desynchronization about 500 ms before the movement onset, however there is an amplitude increase of *beta* rhythms.

Two approaches were followed to extract features for classification.

1) Regression Model: It is proposed a regression model that reflects the BP slow negative shift. Fig. 1b), which evidencies this shift, is clearly fitted with a $2nd$ order regression curve. For the single trial classification, it seems adequate to extract the DC component[1]. Sequences resulting from (1) were initially low-pass filtered by a $8th$ order with 4 Hz cutoff frequency and then fitted with the regression model. The coefficients of the model are obtained through a Least-Squares (LS) estimation. The algorithm is a batch algorithm in that all the data is obtained and then processed in one calculation. The measurement process is modeled as:

$$z = h(t,x) + \varepsilon \tag{2}$$

where (t,z) are pairs of measurements, $h(.,.)$ is the regression curve and models the sensory system, and ε represents both sensor noise and model error. In order to estimate the best vector parameter $\hat{\mathrm{x}}$, all 151 observations (samples of the EEG sequence) are used to minimize the residual ε. In matricial form, the solution of the LS algorithm is [11]:

$$\hat{\mathrm{x}}_{LS} = (H^T H)^{-1} H^T \mathrm{z} \tag{3}$$

where H denotes the matrix that relates the measurements to the unknowns. The regression curve is then the $2nd$ order polynomial:

$$h(t,x) = x_0 + x_1 t + x_2 t^2 \tag{4}$$

2) Wavelet Analysis: Wavelet analysis is a signal processing technique that provides both temporal and frequency information. The discrete wavelet transform (DWT) implemented through the subband coding algorithm analyzes the signal at different frequency and time resolutions [12]. The original signal is successively decomposed into detailed and approximation components. Fig. 2 depicts the wavelet decomposition process. The original signal is chosen to have the last 128 samples of the EEG sequence, $x[n]$. This sequence, which corresponds to a 100 Hz range, is convolved with a low-pass filter $h(n)$ (scaling filter) and then downsampled by a factor of two. The resultant coefficients are designated $cA1$. The original sequence is also convolved with a high-pass filter $g(n)$ (wavelet filter) and then downsampled by a factor of two. The resultant coefficients are designated $cD1$. At this level 1 the signal is represented by half the points and therefore has half the time resolution. All this process is repeated until it reaches level 5. The resultant coefficients are presented in Fig. 3. The top level plot is the original signal corresponding to the overall trial trial (1). The 0-3.125Hz range coefficients represent a coarse approximation of the signal that reflects the BP shift. Detailed coefficients $cD5$, $cD4$ and $cD3$ include respectively *theta*, *mu* and *beta* rhythms. These coefficients will be used as input to the classifier.

D. Classification

This work presents two classifiers based on linear discriminant functions (LDF) for left/right discrimination. The decision regions are separated by hyperplanes. In the first approach, a FLDA reduces each $d-$dimensional feature train vector to a single measurement [13]. In the second approach, a MSE-LDF computes the hyperplane directly in the space of $d-$dimensional train vectors.

1) Fisher Linear Discriminant Analysis: FLDA constructs a linear dimension reduction from the input vector x to a new feature vector y:

$$\mathrm{y} = \mathrm{W}^T \mathrm{x} \tag{5}$$

[1] As it will be seen later this filtering does not provide better classification accuracy.

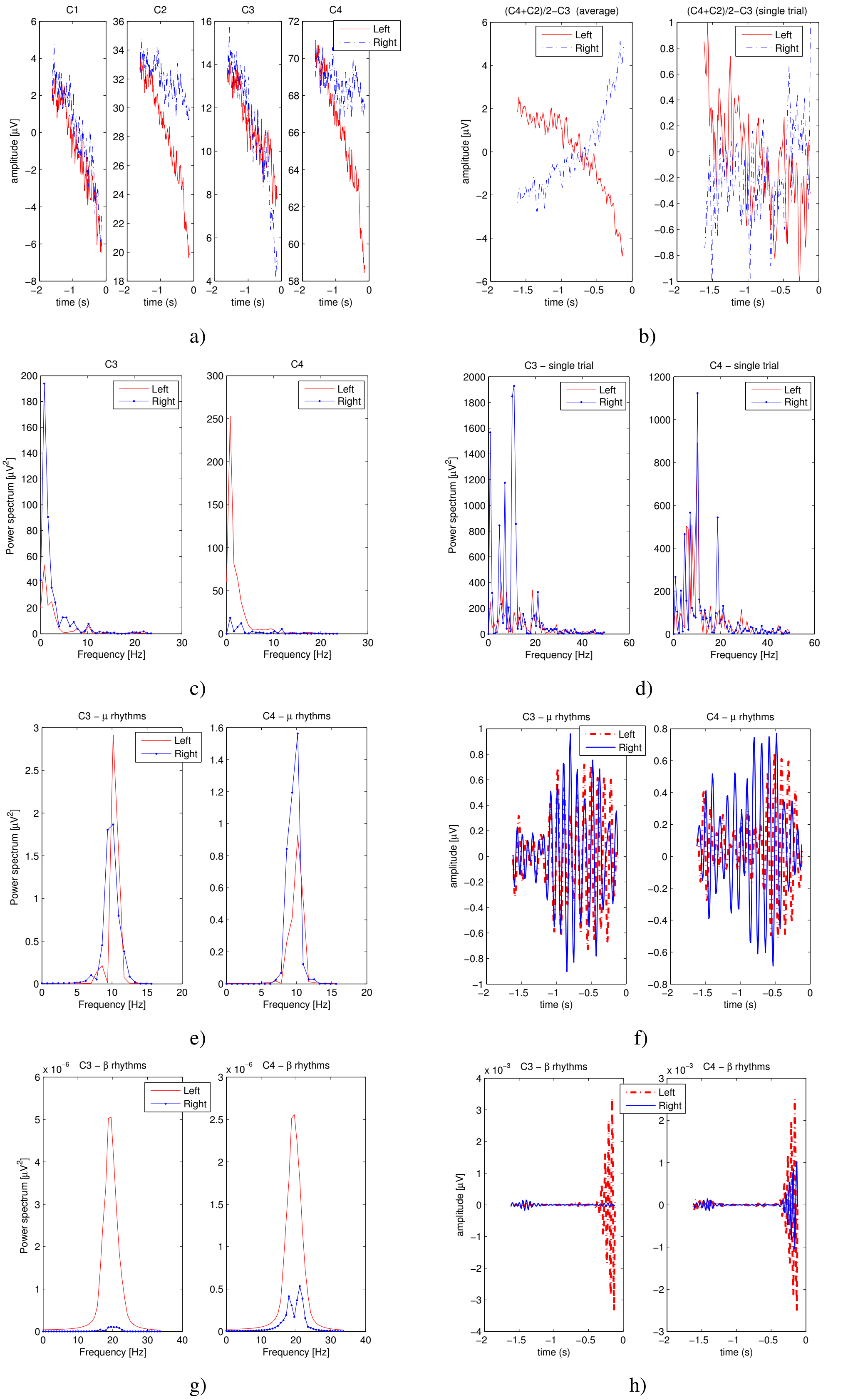

Fig. 1. Average potentials over all trials: a) Scalp in independent channel positions $C1$, $C2$, $C3$ and $C4$ b) In the left, average $\bar{y} = \frac{\bar{C4}+\bar{C2}}{2} - \bar{C3}$; In the right, single trial $y = \frac{C4+C2}{2} - C3$; c) Power spectrum of the average over all trials; d) Power spectrum of a single trial; e) Power spectrum of *mu* rhythms average; f) Time domain output of *mu* band-pass filter; g) Power spectrum of *beta* rhythms average; h) Time domain output of *beta* band-pass filter.

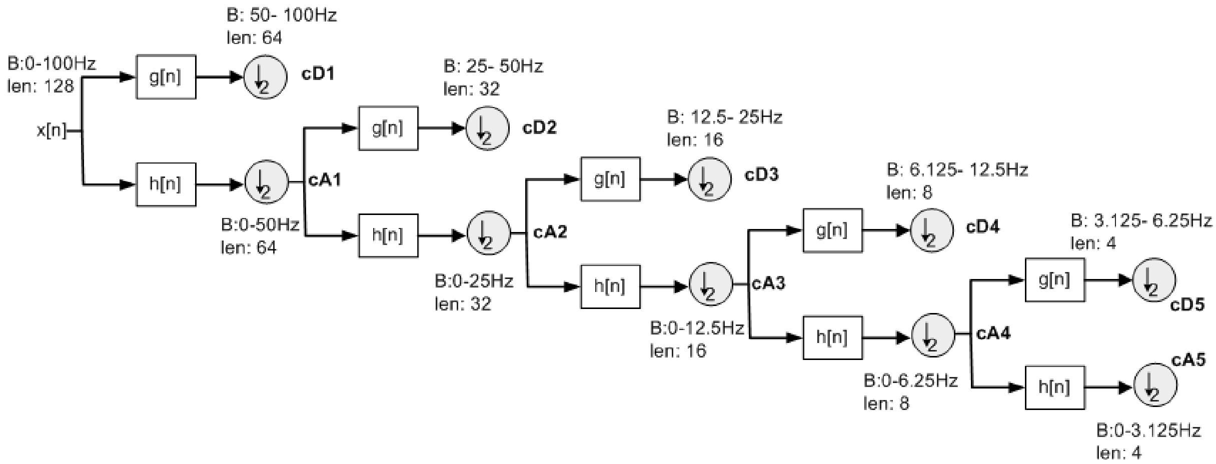

Fig. 2. Wavelet decomposition with subband algorithm

where x $= (x_1|x_2|\cdots|x_n)$ is the vector with all x_i $d-$dimensional samples (training vectors). For the present two class classification case (H_1,H_2), the goal of the FLDA is to maximize the intercluster distance between the two classes and minimize the intracluster within a given class in the new dimension space [13]. Let the within scatter matrix be defined as:

$$S_W = S_1 + S_2 \tag{6}$$

where S_1 and S_2 are the scatter matrices:

$$S_i = \sum_{x \in H_i} (x - m_i)(x - m_i)^T \quad ,i = 1,2 \tag{7}$$

The between scatter-matrix is defined as:

$$S_B = \sum_{i=1}^{2} (m - m_i)(m - m_i)^T \tag{8}$$

where m_i is the mean of the samples in class i, and m is the mean of all samples. The optimal vector **W** is obtained from maximizing the criterion function:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \tag{9}$$

A solution is showed to be:

$$\hat{W} = S_{\mathrm{W}}^{-1}(\mathrm{m}_1 - \mathrm{m}_2) \tag{10}$$

Equation (10) is very similar with Bayesian classifiers in the Gaussian case.

2) Minimum Squared Error: The criterion function of the MSE-LDF involves all $d-$dimensional train vectors. The problem is formulated as finding a solution of a set of linear equations [14]:

$$\mathrm{Y} = \mathrm{w.b} \tag{11}$$

where Y is a $n \times \hat{d}$ $(\hat{d} = d+1)$ matrix that represents the set of $d-$dimensional features vector (*ith* row) and b is the vector target that represents the desired response of the discrimination function. Considering Y as nonsingular, the MSE solution is:

$$\mathrm{w} = (Y^T Y)^{-1} Y^T \mathrm{b} \tag{12}$$

III. Results and Discussion

Labeled data consisting on 219 left events and 194 right events were used for classification training. The 100 event test set was used for testing the accuracy algorithm performance. For correctness purposes and to avoid adjusting parameters that could improve classification performance of the data set, the choice of pre-processing parameters and feature parameters was based without the use of the data set. Notwithstanding this methodology, the results here presented were obtained with the data set, allowing comparison of the results with previous work.

A. Regression

To reduce the dimension of the feature vector for each EEG sequence, a $2nd$ order regression is applied to fit the data. The effect of different low-pass filter cutoff frequencies (*fc*) is analyzed. A $8th$ *Yulewalk* recursive digital IIR low-pass filter is used. The initial cutoff frequency $fc = 4Hz$ was chosen to demonstrate the BP, and the others to analyze the influence of *theta*, *mu* and *beta* rhythms. Contrarily to what it would be expected, best results were not achieved for $fc = 4Hz$ (Table I). This indicates that the slow shift cortical potential in single trials is not readily apparent and therefore has a large fit error. To improve the results, data were fitted with a $3th$ order regression. Results are not conclusive, however best results occurred for $fc = 8Hz$.

The BP is composed by two subcomponents. The first component occurs from 1500 ms to 1000 ms before the onset of the motor movement and is called the *early BP*. The second subcomponent, called *late BP*, occurs about 500 ms before the onset of the movement, and is characterized by a negative shift accentuation. These components are visible in Fig. 1a). Taking advantage of this characteristic, only data from the the last 500 ms (*late BP*) were used for classification. The results improved, reaching a 96% classification accuracy (Table II). Regression fitting of $2nd$ order performed better than $3rd$ order, which indicates that the slow negative shift is now more obvious.

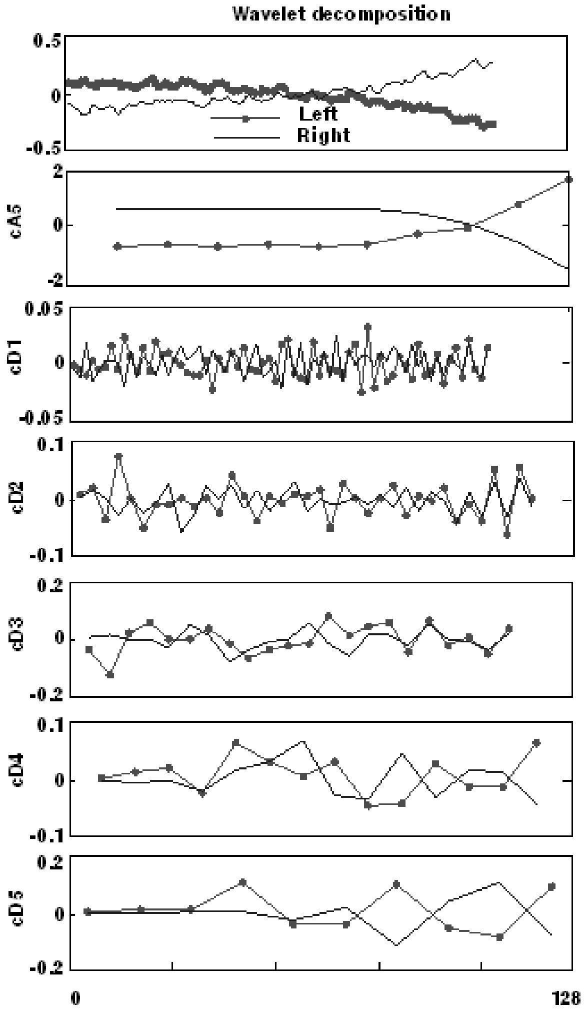

Fig. 3. Coefficients from wavelet decomposition. Original signal has 128 data points from (1) average, $cD1$ 64 points, $cD2$ 32 points, $cD3$ 16 points, $cD4$ 8 points and $cD5$ and $cA5$ 4 points (for all graphs, solid line represents Right and point-solid Left).

TABLE I

CLASSIFICATION ACCURACY WITH REGRESSION FEATURES

	Classifier			
	FLDA		LDF-MSE	
Regression order	2	3	2	3
Low-pass fc				
no filter	91.03%	93.03%	93.16%	94.22%
4.0 Hz	91.15%	92.21%	93.28%	93.28%
8.0 Hz	93.03%	95.04%	92.21%	94.22%
12 Hz	91.03%	94.11%	93.16%	93.16%
22 Hz	91.03%	93.03%	93.16%	94.22%

TABLE II

CLASSIFICATION ACCURACY WITH REGRESSION FEATURES FOR *LATE BP*

	Classifier			
	FLDA		LDF-MSE	
Regression order	2	3	2	3
Low-pass fc				
8.0 Hz	94.10%	95.16%	94.22%	94.22%
22 Hz	96.11%	95.16%	96.11%	96.11%

B. Wavelets

Rather than extract single frequency components, the wavelet decomposition coefficients provides temporal information about frequency bands around the frequencies of interest (Fig. 3). A normal requirement of wavelet analysis is that the wavelet shape look similar to the signal pattern to be localized [7]. Here, the main purpose was to localize a slow shift, therefore the choice fall in Daubechies wavelets (Db2 and Db4). Wavelet decomposition in Fig. 3 was obtained with Db4 (results were slightly better than with Db2). Other wavelets were tested but all with worst classification accuracies. From Fig. 3 it is observed that frequency components above $50Hz$ have very small energy and are constant over all EEG sequence, being considered as background noise. Components within the $25-50Hz$ range ($cD2$) are also not discriminative. Oscillatory *beta* components represented in $cD3$ have more energy, however do not present significant left/right discrimination. Oscillatory components *mu* and *theta* described respectively in $cD4$ and $cD5$ show an obvious discrimination information, specially $cD5$. Coefficients $cA5$ represent an approximation of the original signal, but are better than a low-pass filtered representation of the signal since they provide also data reduction, which improves classification performance. As seen in Fig. 3, for an average trial, the left/right discrimination is clear, and therefore $cA5$ was initially used for classification. The other coefficients were also used as features for classification. Results are presented in Table III. Combining the two better individual features, $cA5$ and $cD5$, an accuracy of 95% was achieved.

C. Comparison

Table IV presents classification accuracies obtained by other researchers following other feature extraction and classification approaches, using the data set made available by [4]. Best results (96.9%) come from the research group who provided the data set. They used the 27 channels which were lowpass filtered at 5 Hz. Features consist on downsampled versions of the original raw EEG sequences, therefore leading to a reduced feature vector dimension. Several classification techniques were tested, namely Fisher Discriminant (FD), Regularized Fisher Discriminant (RFD), Sparse Fisher Discriminant (SFD), Support Vector Machines (SVM) and k-Nearest-Neighbor (k-NN), all achieving very good results. Other authors proposed feature extraction based on stochastic properties [3], common spatial subspace decomposition (CSSD) [15] and wavelet decomposition [16], also with good accuracy results.

In our work a reduced number of recorded channels (C2, C3, C4) was used with a simple normalization preprocessing. The feature extraction, consisting on wavelet decomposition, simultaneously provides low dimension feature vector through the DWT downsampling and temporal information on each significant EEG rhythms as well as respective energy. These characteristics are advantageous over other methods because they provide several feature vectors with low computation time which is suitable for real-time applications. Since the feature vector is low-dimensional,

TABLE IV

ACCURACY OF BCI TECHNIQUES IN PRE-VOLUNTARY MOVEMENT DETECTION USING THE DATASET PROVIDED BY [4]

Preprocessing	EEG Channels	Features	Classification	Accuracy	References
0-5Hz bandpass	motor and somatosensory cortex (21), frontal (5) and occipital (1)	amplitude of downsampled EEG sequences	FD	96.7%	[4]
			RFD	96.9%	
			SFD	96.6%	
			linear SVM	96.4%	
			k-NN	76.9%	
	C2, C3, C4	average and variance	Bayes' rule	92%	[3]
0-3Hz and 8-30Hz bandpass and CSSD	18 channels sensorimotor cortex	output of spatial filter representing ERD and Readiness potential	perceptron	95%	[15]
	C3, CP3, P3, C4, CP4, P4	wavelet coefficients	Gaussian SVM	91%	[16]

good classification results are obtained with simple methods such as FLDA without regularization.

TABLE III

WAVELET FEATURE SELECTION THROUGH CLASSIFICATION ACCURACY

Coefficients	FLDA	LDF-MSE
cA5	94.10%	93.16%
cD2	53.47%	54.78%
cD3	68.00%	68.00%
cD4	73.20%	73.20%
cD5	93.16%	93.16%
cA5&cD5	95.16%	95.16%

IV. CONCLUSION

Our single trial classification of EEG sequences for left/right self-paced tapping discrimination has been successful. Both regression fitting and wavelet decomposition proved to be suitable feature extraction techniques. It was obtained a 96% and 95% classification accuracy for regression and wavelet methods, respectively. The FLDA and LDF-MSE linear classifiers had similar performance. The good results are encouraging, however the detection of upcoming events without any synchronization cue is essential for an effective BCI. The authors propose as future work to investigate this topic. The temporal properties of the wavelets can give an important contribution for asynchronous BCI.

V. ACKNOWLEDGMENTS

This work has been in part supported by Fundação para a Ciência e a Tecnologia (FCT), under Grant POSC/EEASRI/58279/2004. Gabriel Pires would like also to thank the support of FCT through the research fellowship SFRH/BD/29610/2006.

REFERENCES

[1] B. Blankertz, K. Muller, G. Curio, G. Vaughan, T. Schalk, J. Wolpaw, A. Schlogl, C. Neuper, G. Pturtscheller, T. Hinterberger, M. Schroder, 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, June 2004.

[2] N. Neumann and N. Birbaumer. Predictors of successful self control during brain-computer communication. *J. Neurol. Neurosurg. Psychiatry*, 74:1117–1121, 2003.

[3] J. Kohlmorgen and B. Blankertz. Bayesian classification of single-trial event-related potentials in EEG. *Int. J. Bif. Chaos*, 14(2):719–726, 2004.

[4] B. Blankertz, G. Curio, and K. Muller. Classifying single-trial EEG: Towards brain computer interfacing. *Advances in Neural Information Processing Systems 14 NIPS01. MIT Press*, pages 157–164, 2002.

[5] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao, and F. Yang. BCI competition 2003-data set IV: An algorithm based on CSSD and FDA for classifying single-trial EEG. *IEEE Transactions on Biomedical Engineering*, 51(6):1081–1086, June 2004.

[6] G. Dornhege, B. Blankertz, G. Curio, and K. Muller. Boosting bit rates in non-invasive EEG single-trial classification by feature combination and multi-class paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6):993–1002, 2004.

[7] Glassman E. A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs. *IEEE Transactions on Biomedical Engineering*, 52(11):1851–1862, November 2005.

[8] B. Blankertz, G. Curio, and K. Muller. EEG self-paced key typing dataset. *NIPS 2001 Brain Computer Interface Post Workshop Data Competition. Online available: http://liinc.bme.columbia.edu/competition.htm*.

[9] R. Q. Cui, D. Huter, and L. Deecke. Neuroimage of voluntary movement: Topography of the bereitschaftspotential, a 64-channel DC current source study. *Neuroimage*, 9:124–134, 1999.

[10] N. Birbaumer, H. Thomas, A. Canavan, and R. Brigitte. Slow potentials of the cerebral cortex and behavior. *The American Physiologic Society*, 70(1):1–41, January 1990.

[11] F. Heijden, R. Duin, D. Ridder, and D. Tax. *Classification, Parameter Estimation and State Estimation*. John Wiley & Sons, 2004.

[12] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[13] R. Schalkoff. *Pattern Recognition, Statistical, Structural and Neural Approaches*. John Wiley & Sons, 1992.

[14] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley London UK, 2001.

[15] Y. Li, X. Gao, H. Liu, and S. Gao. Classification of single-trial electroencephalogram during finger movement. *IEEE Transactions on Biomedical Engineering*, 51(6):1019–1025, June 2004.

[16] A. Yong, G.C.M. Silvestre, and N.J. Hurley. Single-trial EEG classification for brain-computer interface using wavelet decomposition. *EUSIPCO conference*, September 2005.