A Motor Imagery BCI Experiment using Wavelet Analysis and Spatial Patterns Feature Extraction

Obed Carrera-León[1], Juan Manuel Ramirez[1], Vicente Alarcon-Aquino[2], Mary Baker[3], David D´Croz-Baron[1], Pilar Gomez-Gil[4]

[1]Electronics Departments, INAOE, Tonantzintla, Puebla, Mexico
[2]Department of Computing, Electronics and Mechatronics, UDLAP, Puebla, Mexico
[3]Department of Electrical Engineering, Texas Tech University, Lubbock, Texas, USA
[4]Computer Science Department, INAOE, Tonantzintla, Puebla, Mexico

Abstract - A brain computer interface (BCI) is a system that aims to control devices by analyzing brain signals patterns. In this work, a convenient time-frequency representation (TFR) for visualizing ERD/ERS phenomenon (Event related synchronization and desynchronization) based on Hilbert transform and spatial patterns is addressed, and a wavelet based feature extraction method for motor imagery tasks is presented. The feature vectors are constructed with four statistical and energy parameters obtained from wavelet decomposition, based on the sub-band coding algorithm. Experimentation with three classification methods for comparison purposes was carried out using Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Support Vector Machine (SVM). In each case, ten-fold validation is used to obtain average misclassification rates.

Keywords - Wavelet (DWT), Support Vector Machine, LDA, BCI, EEG.

I. INTRODUCTION

Human brain is the most complex structure that exists in the human body. In the last years, technological advances have allowed measuring the electric and magnetic fields generated by brain neurons, which are the cell structures in charge of the communication within the brain [1]. The brain activity can be analyzed through an Electroencephalograph (EEG) that shows the post-synaptic potentials produced inside the brain reflected to the scalp [2]. A brain-computer interface (BCI) is a term broadly used to describe a system which translates the electrical signals generated from cognitive processes into control signals for a variety of applications, such as computer controls, speech synthesizers, or mechanical prostheses. Fig. 1 shows a typical block diagram of a brain-computer interface system, describing the several steps involved. This work concentrates on the data signal processing steps located inside of the box. The ongoing electroencephalographic signals (EEG) contain information associated to movements, mental tasks or mental responses related to some stimuli. These signals are analyzed and processed through several mathematical techniques to extract useful information represented in the form of feature vectors, which are then translated into meaningful control commands [1-3]. Currently, BCI systems can be classified in seven categories, according to the neuro-mechanism used: sensory motor activity, slow cortical potentials (SCP), P300, visual evoked potentials (VEP), response to mental tasks, activity in neurons cells and combinations of the above [4-6]. The task addressed in this paper corresponds to motor imagery (MI) obtained from the sensory motor neuro-mechanism. In general, two types of patterns are usually present in this activity: event related potentials (ERP), detected as energy changes in μ and β bands generated when a voluntary movement is performed, and movement related potentials (MRP), which are low frequency patterns that appears between 1 – 1.5 s before movement. In the first case, the event related potentials consist, in general terms, in decrements or increments of the energy on the ongoing EEG signal at certain frequency bands, which are described in the literature as the ERD/ERS phenomenon (Event Related Desynchronization and Synchronization) [7,8].

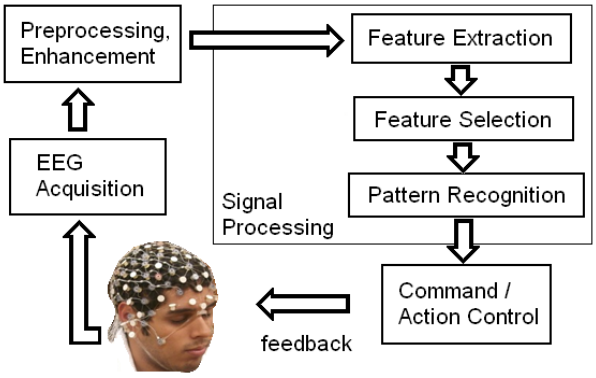


Figure 1. Block diagram of a brain-computer-interface

The main issue involved in the motor imagery pattern recognition process is to successfully estimate, visualize and represent the ERD/ERS phenomenon in a feature vector. Several feature extraction techniques have been used in MI, such as: amplitude values of EEG [9], band power [10], power spectral density [11,12], auto-regressive (AR) and adaptive auto-regressive models (AAR) [13], windowed and fast Fourier analysis, cross correlation, and some others. As these ERPs are locked in time but not in phase and they are highly non-stationary [14,18-20], the detection of these patterns turns into a difficult task in which time-frequency techniques could provide significant information. Thus, based on wavelet theory and in recent studies, wavelet analysis provides an excellent tool for extracting significant information about the time-frequency behavior of the signal, with a good tradeoff on performance and computational complexity. In order to

978-1-4673-0870-0/12/$31.00 ©2012 IEEE

identify and quantitatively represent ERD/ERS phenomena, several methods have been proposed such as windowed Fourier analysis, Hidden Markov models and continuous wavelet transform [19], among others. In this paper we propose a feature extraction method based on Hilbert transform and band pass filtering of the EEG signals, aiming to feature extraction of ERD/ERS phenomena on a motor imagery experiment using spatial patterns. Furthermore, the Discrete Wavelet Transform (DWT) is a transformation that can be used to analyze the temporal and spectral properties of non-stationary signals, with a very good localization in time and frequency, which is one of the most attractive features of wavelet analysis. DWT is an operation limited in scale according to the sampling frequency and the number of decomposition levels. A feature extraction method based on statistical information obtained from wavelet analysis, is used in this work as the input of three different classifiers: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Support Vector Machine (SVM). A comparison on performance of these approaches is presented in the results section.

II. METHODOLOGY

A. Time-frequency representation

The common ERD/ERS detection involves four steps, [19, 20] which includes: band-pass filtering of all event-related trials, squaring of the amplitude samples, averaging of power samples across all trials and averaging over time samples to smooth the data and reduce the variability. In order to visualize the ERD/ERS phenomena a TFR was constructed with a set of band pass filters and the Hilbert transform.

A 5th order Butterworth filter bank was used to separate the frequency bands. The coefficients of this array are calculated for a variable frequency range (4-30 Hz) and with bandwidths from 1 Hz to 4 Hz with 50% of overlapping that guarantee an increase in the information redundancy of neighboring bins [12-14]. A narrow band filtered signal can be seen as an amplitude-modulated signal expressed as:

$$g(t) = a(t)\cos(2\pi f_0 t + \theta) \quad , \tag{1}$$

where $g(t)$ contains frequency information of time locked ERD/ERS at band f_0 with initial phase θ [12,13]. The Hilbert transform of (1) can be obtained as [16]:

$$\hat{g}(t) = a(t)\sin(2\pi f_0 t + \theta) \tag{2}$$

and the analytic signal as:

$$z(t) = g(t) + j\hat{g}(t) \quad , \tag{3}$$

where the envelope is extracted as follows:

$$w(t) = |z(t)| = \sqrt{g^2(t) + j\hat{g}^2(t)} \tag{4}$$

Fig. 2 shows the filterbank distribution. It can be noticed the flat response in the band pass section and a slow but remarked decay, which give some redundancy to the TFR. The envelope extraction process applied to one frequency band is illustrated in Fig. 3. An ERD obtained through averaging over the set of trials appears right after the event cue [19-21]. Fig. 4 shows an example of the reactive frequency which spread around 10 Hz where ERD/ERS remarkably appears over time. It can be seen the differences in the time course of the event before and after the event cue.

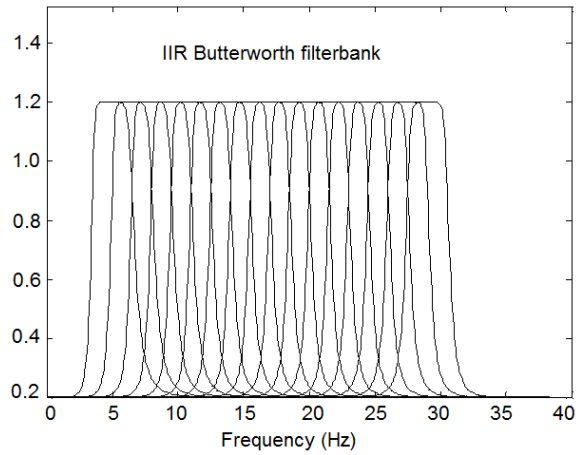
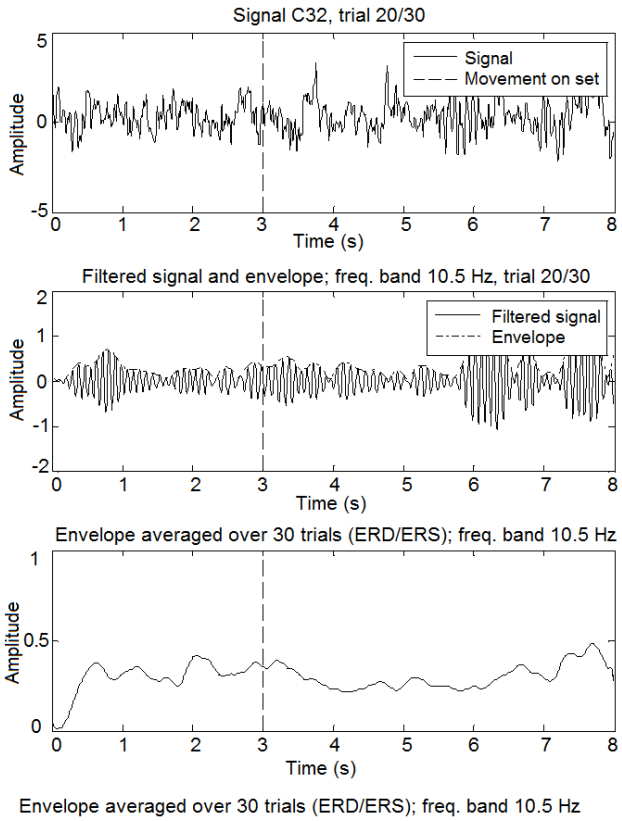


Figure 2. Butterworth fifth order filterbank with BW = 3Hz from 5 to 30 Hz.



Envelope averaged over 30 trials (ERD/ERS); freq. band 10.5 Hz

Figure 3. From top to bottom: raw EEG signal, filtered signal with envelope averaged over the set of trials.

Fig. 5 and 6 show typical obtained reference spatial patterns obtained during a trial, before and after feature reduction. It is evident the energy change registered after the event cue.

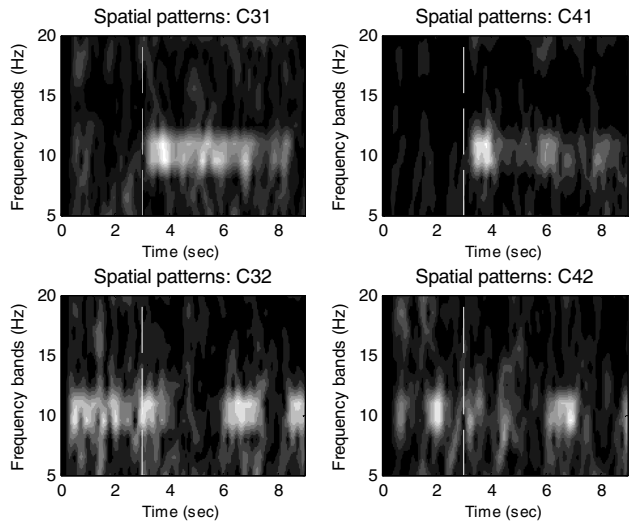


Figure 4. Single trial TFR (5 – 20 Hz). The event cue is indicated by a white dotted line.

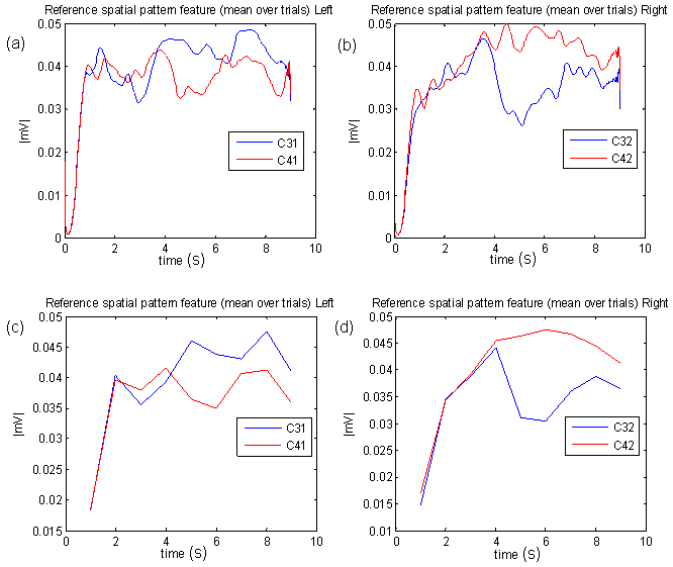


Figure 5. Reference spatial pattern features corresponding to left and right electrodes before feature reduction (a,b), and after feature reduction (c,d).

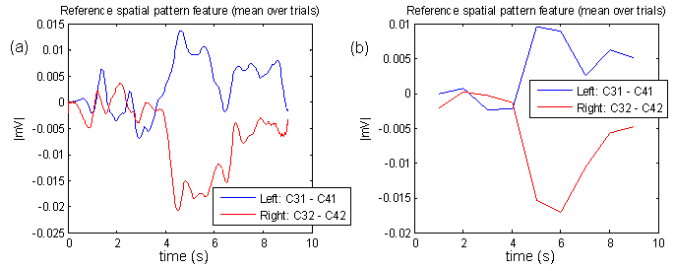


Figure 6. Differenced reference spatial pattern features before feature reduction (a) and after feature reduction (b).

B. *Discrete wavelet transform*

DWT gives precise time-frequency information about the signal. It decomposes the signal in a number of sub-bands at different scales, according to the number of decomposition levels. DWT is defined by two functions referred as scaling and detail functions [21-26].

$$\varphi(t) = \sqrt{2}\sum_{l\in\mathbb{Z}} h_0\varphi(2t-l), \tag{5}$$

$$\psi(t) = \sqrt{2}\sum_{l\in\mathbb{Z}} h_1\varphi(2t-l). \tag{6}$$

The frequency responses of equation 5 and 6 correspond to low pass and high pass FIR filters, respectively; h_0 are the low-pass filter coefficients and h_1 are the high-pass filter coefficients related to a chosen wavelet φ. From these equations a signal can be estimated as:

$$x(t) = \sum_{k\in\mathbb{Z}} c_{J,k}\varphi_{J,k}(t) + \sum_{j=J}^{\infty}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(t), \tag{7}$$

where

$$c_{j,k} = \sum_m h_0 c_{j-1,2k-l} \tag{8}$$

$$d_{j,k} = \sum_m h_1 c_{j-1,2k-l}. \tag{9}$$

Approximation and details coefficients are obtained through equations 8 and 9, which describe a set of recursive quadrature mirror filters. The so-called sub-band coding algorithm is illustrated in Fig. 7 with a two wavelet decomposition levels. The frequency ranges corresponding to each sub-band depends on the sampling frequency of the signal. Table 1 shows the frequency distribution for a three level DWT and a 128 Hz sampling frequency.

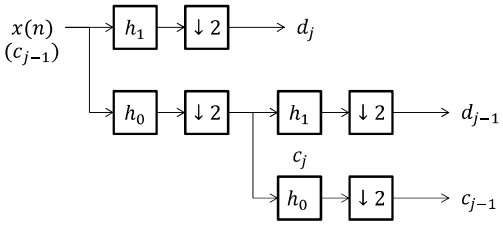


Figure 7. Two stage filter bank decomposition tree.

Table 1. Frequency distribution of a four level DWT. D = detail, A = approximation for Fs = 128 and 250 Hz.

Decomposition level	*Frequency range (Hz)*
D1	32-64
D2	16-32
D3	8-16
A3	0-8

C. *Support Vector Machines (SVM)*

SVMs have been proposed for pattern recognition in a wide range of applications by its ability for learning from experimental data, and its effectiveness over some other conventional parametric classifiers. Briefly, SVM is a statistical learning method based on a structural risk minimization procedure, which minimizes the upper bound of the generalization errors consisting of the sum of training errors and a confidence interval [26]. The original input space is transformed through a non-linear feature mapping to a high dimensional feature space, where the data is linearly separable by a hyperplane. The goal during the training process is to find the separating hyperplane with the largest margin in the obtained hyperspace. The transformation is performed using a non-linear function referred as the transformation kernel. There are three common kernels used for the non-linear feature mapping: polynomial, radial basis function, and sigmoid kernels. Linear hyperplane classifiers are based on the class of decision functions [26,27]. The optimal hyperplane is defined as the one with the maximal margin of separation between the two classes. The solution of a constrained quadratic optimization process can be expanded in terms of a subset of the training patterns called support vectors that lie on the margin:

$$w = \sum_{i=1}^{N} v_i x_i \tag{19}$$

Thus the decision rule depends only on dot products between patterns:

$$y = sign\left(\sum_{i=1}^{N} v_i (x_i \cdot x) + b \right) \tag{20}$$

The above linear algorithm is performed in the new feature space obtained through some non-linear transformation Ø by using some of the described kernels. The kernel is related to the Ø function by:

$$K(x_i, x) = \phi(x_i) \cdot \phi(x) \tag{21}$$

In this work the radial basis function (RBF) was used. RBF is defined as:

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / \sigma} \tag{22}$$

Classification of a test sample x is then performed by:

$$y = sign\left(\sum_{i=1}^{N} \alpha_i v_i K(x_i, x) + b \right), \tag{23}$$

where N is the number of training samples, v_i is the class label, α_i a Lagrangian multiplier, the elements x_i for which $\alpha_i > 0$ are the support vectors, and $K(x_i, x)$ is the function kernel.

III. DESCRIPTION OF THE EXPERIMENT

A. *Experimental paradigm and databases*

The EEG data used in this work was obtained from two sources: 1) the public repository of the BCI Competition IIIB, available to the international community for academic and research purposes, and 2) an own database generated according to the experimental paradigm of the mentioned competitions, built in the Autumn's Dawn NICE (Neuro-Imaging Cognition and Engineering) Laboratory at Texas Tech University. Fig. 8 shows the experimental setup. A detailed description of the signal acquisition and experimental paradigm can be found in [6]. The EEG was obtained using a sampling frequency of 125 Hz, and pass-band filtered between 0.5 and 30Hz. Each trial started with the presentation of a fixed cross at the center of the monitor, followed by a short warning tone at 2s. At 3s, the fixed cross was overlaid with an arrow at the center of the monitor for 1.25 s, pointing either to the left or to the right as experimental cue. Depending on the direction of the arrow, the subject was instructed to imagine a movement of the left or the right hand. Fig. 9 shows the trials structure. The experiment was implemented using 64 electrodes HCGSN from EGI, using a standard 10-10 distribution, as shown in Fig. 10.

Figure 8. Experimental EEG setup used in the experiments

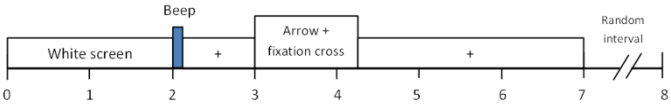


Figure 9. Structure of the trials used during the motor imagery experiments.

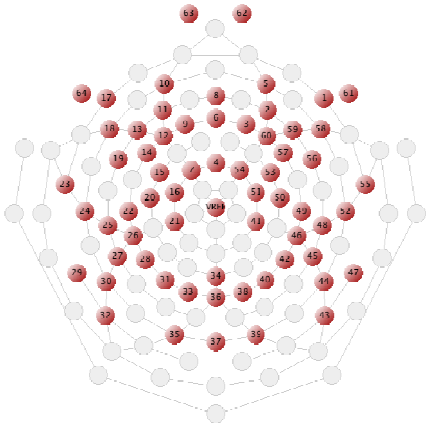


Figure 10. Standard 10-10 electrodes distribution corresponding to the 64 electrodes EGI system.

According to the described paradigm, the generated database corresponded to EEG experiments with imagined left and right movements for two subjects. The EEG was obtained using a sample frequency of 125Hz, and pass-band filtered between 0.3 and 30Hz. The trial duration was 7 seconds, and in each trial the subject was instructed to imagine only one movement, depending on the direction of the arrow at time t = 3s. Data set 1 provides the information of electrodes C3, Cz and C4. Signal Cz was further subtracted from C3 and C4 in order to improve the signal to noise ratio.

B. Feature extraction

For each electrode C3 and C4 a MxN array was created using i trials, with $i = 1,2,3,\dots,M$ and $\frac{N}{Fs}$ seconds per trial. Three-level wavelet decomposition was applied to each trial. After signal decomposition, 2 statistical features were extracted aiming to the feature vector generation: sub-band average power and standard deviation of each sub-band coefficients set. The feature vector was then constructed according to the following equation:

$$WF_i^{Cec} = \left[\frac{1}{m}\sum_{n=1}^{m}(D3(n))^2, \frac{1}{m}\sum_{n=1}^{m}(D2(n))^2, std(D3), std(D2)\right], \quad (10)$$

where e is the electrode number, $c = 1,2$ is the class, and m is the number of coefficients per sub-band. The signal was analyzed with a $2\,s$ sliding rectangular window for $3 < s < length(x_i(t))$, according to the event cue. The window was slide at every sampling point, and a feature vector was obtained for each shift. The classification rate was estimated by a ten-fold-cross-validation using Fisher LDA, SVM and QDA.

For the SVM, a Radial Basis Function was used as kernel with $\sigma = 1$ described by [15]:

$$k(x, x') = \exp\left(-\left|\left|x - x'\right|\right|^2/(2 * \sigma^2)\right) \quad (11)$$

The SVM was implemented using the SVM toolbox for Matlab developed by Steve Gunn, from Image Speech and Intelligent Systems Group, University of Southampton.

IV. Results

Table 2 shows the results obtained from the motor imagery experiments using a feature extraction method based on spatial patterns and Hilbert transform, for 9 different subjects. The results are presented in each case as a miss-classification rate (MCR). The database was divided into training and testing trials in order to perform a ten-fold cross-validation. Once the classifier is trained with the corresponding training set, the MCR is obtained by introducing one by one all of the testing trails, and averaging the partial results.

Table 2. Comparison on MCR obtained using the three classification methods based on spatial patterns

Subject	*FLDA*	*QDA*	*SVM*
S1	0.1643	0.2357	0.2786
S2	0.2357	0.4143	0.4643
S3	0.3643	0.4571	0.5214
S4	0.3643	0.5143	0.5286
S5	0.2467	0.2400	0.2500
S6	0.5143	0.4754	0.4795
S7	0.4403	0.4925	0.4571
S8	0.4813	0.4944	0.4552
S9	0.5112	0.5056	0.5336
Average MCR	0.36916	0.42548	0.44092

Table 3 shows the averaged results obtained through the experiments using a feature extraction method based on wavelet analysis for 11 different wavelet functions (db3 – db5, coiflet3 – coiflet5, sym3 – sym7). These results were obtained using ten-fold validation. The best results were obtained using Coiflet-4 and Daubechies-5 wavelets.

Table 3. Comparison on MCR obtained using the three classification methods based on wavelet analysis

Wavelet	*LDA*	*QDA*	*SVM*
coif2	0.1286	0.1500	0.2332
coif3	0.1286	0.1571	0.2424
coif4	0.1214	0.1286	0.2464
db3	0.1286	0.1643	0.2183
db4	0.1214	0.1357	0.2361
db5	0.1286	0.1286	0.2218
sym3	0.1286	0.1643	0.2258
sym4	0.1357	0.1500	0.2251
sym5	0.1357	0.1500	0.2319
sym6	0.1357	0.1571	0.2278
sym7	0.1286	0.1571	0.2335

V. Conclusions

In this work, a series of experiments oriented to the detection of the ERD/ERS phenomenon over EEG motor imagery were presented. Feature extraction process was carried out using two methodologies: statistical features of wavelet coefficients, and spatial patterns based on Hilbert transform. In average, feature extraction based on wavelet analysis provided better results. A comparison on the miss-classification rate obtained using a series of wavelets showed

that best results were obtained using Coiflet-4 and Daubechies-5 wavelet type. Further experiments incorporating a feature selection process are currently in progress.

ACKNOWLEDGMENTS

The first author gratefully acknowledges the financial support from the Mexican National Council for Science and Technology (CONACYT), and the Autumn's Dawn NICE (Neuro-Imaging Cognition and Engineering) Laboratory at Texas Tech University for the research stay. This research is supported by CONACYT Grant No. CB-2010-155250.

VI. REFERENCES

[1] F. Lotte, M. Congedo, A. L'ecuyer, F. Lamarche, B. Arnaldi "A review of classification algorithms for EEG-based brain–computer interfaces", in J. Neural Eng. Vol. 4, 2007, pp. R1–R13.

[2] J. B. Fisch, R. Spehlmann, *EEG primer: basic principles of digital and analog EEG,* Elsevier, 1999.

[3] A. Bashashati, M. Fatourechi, R.K. Ward, G.E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals", in J. Neural Eng., Vol. 4, 2007, pp. R32–R57.

[4] P.W. Ferrez, J.R. Millan, "Simultaneous real time detection of motor imagery and error-related potentials for improved BCI accuracy", in 4th International Brain-Computer Interface Workshop and Training Course, Graz, Austria, 2008.

[5] B. Z. Allison, J.A. Pineda, "ERPs evoked by different matrix sizes: implications for a brain–computer interface (BCI) system", in IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 11, 2003, pp. 110-113.

[6] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, H. Ritter "BCI competition 2003–data set IIIb: support vector machines for the p300 speller paradigm", in IEEE Trans. Biomed. Eng. Vol. 51, 2004, pp. 1073–1076.

[7] G. Pfurtscheller, "Spatiotemporal ERD/ERS patterns during voluntary movement and motor imagery", in Supplements to Clinical Neurophysiology, Vol. 53, 2000, pp 196-198.

[8] S. Chiappa, S. Bengio S, "HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems", in European Symposium on Artificial Neural Networks ESANN, 2004.

[9] J.R. Millan, J. Mouriño, "Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project", in IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 11, 2003, pp. 159-61.

[10] G. Pfurtscheller, C. Neuper, A. Schlogl, K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters", in IEEE Trans. Rehabil. Eng., Vol. 6, 1998, pp. 316–325.

[11] G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement", in Electroenceph. Clin. Neurophysiol. Vol. 103, 1997, pp. 642–651.

[12] T. Wang, B. He, "An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain–computer interface", in J. Neural Eng. Vol. 1, 2004, pp. 1–7.

[13] T. Wang, J. Denga, B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns", in Clinical Neurophysiology, Vol. 115, 2004, pp. 2744–2753.

[14] P. Durka, *Matching Persuit and Unification in EEG Analisys,* Artech House, Inc. Norwood, MA, 2007.

[15] F.R. Kschischang, *The Hilbert Transform,* The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 2006.

[16] S.A. Mallat, *A wavelet tour of signal processing: the sparse way*, Academic Press, 2009.

[17] B. Graimann, G. Pfurtscheller, "Quantification and visualization of event-related changes in oscillatory brain activity in the time–frequency domain", in Progress in Brain Research, Vol. 159, 2006, pp 79-97.

[18] G. Pfurtscheller, F.H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", in Clinical Neurophysiology, Vol. 110, 1999, pp. 1842-1857.

[19] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model", in Expert Systems with Applications, Vol. 32, 2007, pp. 1084–1093.

[20] V. Alarcon-Aquino, *Anomaly Detection and Prediction in Communication Networks Using Wavelet Transforms,* PhD Thesis, Imperial College London, London UK, 2003.

[21] A. Subasi, "Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction", in Computers in Biology and Medicine, Vol. 37, 2007, pp. 227–244.

[22] A. Kandaswamy, C.S. Kumar, R.P. Ramanathan, S. Jayaraman, N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients", in Computers in Biology and Medicine, Vol. 34, No. 6, 2004, pp. 523–537.

[23] E.D. Ubeyli, "Probabilistic neural networks combined with wavelet coefficients for analysis of electroencephalogram signals", in Expert Systems, Vol. 26, No. 2, 2009, pp. 147-159.

[24] Q. Xu, H. Zhou, Y. Wang, J. Huang, "Fuzzy support vector machine for classification of EEG signals using wavelet-based features", in Medical Engineering & Physics, Vol. 31, 2009, pp. 858–865.

[25] S. Walker, *A Primer on Wavelets and their Scientific Applications, Second Edition,* Chapman and Hall, CRC, 2008.

[26] B. Scholkopf, A.J. Smola, *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond,* MIT Press. London, England, 2002.

[27] N. Cristianini, J. Shawe-Taylor, *An introduction to Support Vector Machines and Other Kernel-based Learning Methods,* Cambridge University Press, 2000.