Biomedical Engineering: Applications, Basis and Communications, Vol. 25, No. 6 (2013) 1350058 (13 pages)
DOI: 10.4015/S1016237213500580

FEATURES EXTRACTION METHOD FOR BRAIN-MACHINE COMMUNICATION BASED ON THE EMPIRICAL MODE DECOMPOSITION

Pablo F. Diez[*,†,§], Vicente A. Mut[†], Eric Laciar[*],
Abel Torres[‡] and Enrique M. Avila Perona[†]

[*]*Gabinete de Tecnología Médica*
Universidad Nacional de San Juan
Av. San Martín 1109 (oeste)
Capital, San Juan (J5400ARL)
San Juan, Argentina

[†]*Instituto de Automática*
Universidad Nacional de San Juan
San Juan, Argentina

[‡]*Department ESAII*
Universitat Politècnica de Catalunya
Institut de Bioenginyeria de Catalunya (IBEC)
and CIBER de Bioingeniería, Biomateriales y
Nanomedicina (CIBER-BBN), Barcelona, Spain

[§]*pdiez@gateme.unsj.edu.ar*

Accepted 1 January 2013
Published 24 July 2013

ABSTRACT

A brain-machine interface (BMI) is a communication system that translates human brain activity into commands, and then these commands are conveyed to a machine or a computer. It is proposes a technique for features extraction from electroencephalographic (EEG) signals and afterward, their classification on different mental tasks. The empirical mode decomposition (EMD) is a method capable of processing non-stationary and nonlinear signals, as the EEG. The EMD was applied on EEG signals of seven subjects performing five mental tasks. Six features were computed, namely, root mean square (RMS), variance, Shannon entropy, Lempel−Ziv complexity value, and central and maximum frequencies. In order to reduce the dimensionality of the feature vector, the Wilks' lambda (WL) parameter was used for the selection of the most important variables. The classification of mental tasks was performed using linear discriminant analysis (LDA) and neural networks (NN). Using this method, the average classification over all subjects in database is $91 \pm 5\%$ and $87 \pm 5\%$ using LDA and NN, respectively. Bit rate was ranging from 0.24 bits/trial up to 0.84 bits/trial. The proposed method allows achieving higher performances in the classification of mental tasks than other traditional methods using the same database. This represents an improvement in the brain-machine communication system.

Keywords: Brain-machine interface (BMI); Brain-computer interface (BCI); Empirical mode decomposition (EMD); Feature extraction.

[§]Corresponding author: Pablo F. Diez, Gabinete de Tecnología Médica, Universidad Nacional de San Juan, Av. San Martín 1109 (oeste), Capital, San Juan (J5400ARL), San Juan, Argentina,Tel: (+54)0264-4211-700 ext. 313; Fax: (+54)0264-4200-289; E-mail: pdiez@gateme.unsjedu.ar

INTRODUCTION

Some diseases, such as stroke, affect the communication between the brain and the body of people. Such diseases prevent them to move, to speak or to perform other daily activities. Consequently, those people cannot (partially or totally) communicate with their relatives or with the surrounding environment. Hence, a different communication way must be provided to people with disabilities.

A brain-machine interface (BMI), or specifically a brain-computer interface (BCI), is a system that provides an alternative channel of communication between the brain and the environment around an individual having neuromuscular disabilities.[1] A principal stage of a BCI system is the features extraction. This obtains relevant features from brain signals related to the intention of the user.

Generally, the brain activity is obtained using electroencephalography (EEG) due to low costs and non-invasiveness. The EEG measures the activity of many neurons (attenuated by thick layer of tissue, such as fluids, bones and skin) and corrupting random noise.[2] Different signal processing methods have been proposed to extract features from EEG signals.

A review of signal processing techniques in feature extraction on EEG signal used in the BCI field is shown in McFarland *et al.*[3] Involving the traditional signal processing techniques, an extended utilized method is Fourier transform (FT); which is a helpful technique for the analysis of linear and stationary signals. Another method for signal processing is the Wavelet transform (WT). For the current case, this analysis is valuable for linear and non-stationary signals. In practice, these methods are useful when assuming that the EEG signal is linear and stationary (FT case), or non-stationary (WT case) for short time periods. Nevertheless, the EEG signal is essentially nonlinear and non-stationary, therefore these processing techniques may not be capable of extracting all the features of EEG signals. The empirical mode decomposition (EMD) technique has the unique property of dealing with nonlinear and non-stationary time series.[4,5] Continuous WT and EMD are both analysis methods in time-frequency domain. The EMD acts essentially as a filter bank resembling those involved in wavelet decompositions. However, in wavelet analysis this sub-band filtering is pre-determined and, on the contrary, in EMD this sub-band filtering is signal dependent. Details about EMD-wavelet differences are described in Huang *et al.*[4] Hence, EMD is proposed as a tool for processing and extracting features of EEG signals.

EMD is a relatively new technique and it is currently employed in many fields, one of which is biomedical engineering, with a wide acceptance. Indeed, there have been several applications using EMD, e.g. in ECG de-noising,[6] for analyzing neural data,[7] for analyzing respiratory mechanomyographic signals,[8] and involving EEG signals processing, for analyzing hypoxia EEG[9] and epileptic seizure detection[10] as well.

In the BCI area, EMD has been used to analyze steady-state evoked potential,[11] P300 evoked potentials[12] and to extract features from motor-imagery. For example, EMD was used to detect α-rhythm desynchronization[13] and post-movement beta activities from 32-channel EEG data when performing a right index finger movement.[14] Besides, EMD was used to study the active frequency range that corresponds with the motor imagery of each subject.[15] Detection of α-rhythm and β-rhythm with combination of EMD and wavelet packet transform is proposed in Yuan *et al.*[16] Recently, EMD and asymmetries of brain activity was used to detect mental tasks.[17]

In this work, the EMD is employed to extract features from ongoing EEG, aimed at controlling a future BCI. Preliminary results of the proposed method were already published.[18] The proposed method achieves high accuracies in the classification of mental tasks. Moreover, the obtained results are higher than reported in related bibliography and evaluated on the same database.[34–36]

MATERIALS AND METHODS

Materials

The EEG database utilized in this work was made by Keirn and Aunon[19] and is available online.[20] The subjects were seated comfortably in a sound controlled booth with dim lighting. Electrodes were placed at O_1, O_2, P_3, P_4, C_3 and C_4 and referenced to electrically linked mastoids, A_1 and A_2. The electrodes were connected through a bank of amplifiers (Grass7P511), whose band-pass analogical filters were set at 0.1−100 Hz. The data were sampled at 250 Hz with a Lab Master 12-bit A/D converter. Signals were recorded for 10 sec during each mental task and each task was repeated for 10 sessions. Seven subjects, 21 to 48 years old, participated in the study involving a total of five distinct tasks. The subjects were instructed not to vocalize or make overt movements while solving the mental tasks:

Baseline Task (*Base*): There was no mental task to be performed here. The subject was told to simply relax and try to think of nothing in particular.

Mathematical Multiplication Task (*Math*): The subject was given a non-trivial multiplication problem to solve. The problems were non-repeating and were designed to prevent the subject from completing the task before the end of the 10-sec recording session.

Geometric Figure Rotation (*Rot*): A three-dimension block figure was shown during 30 sec to the subject, after which the drawing was removed and the subject was instructed to visualize the object being rotated about an axis. The EEG was recorded during the mental rotation period.

Mental Letter Composing (*Lett*): The subject was instructed to mentally compose a letter to a friend or relative without vocalizing. Since the task was repeated several times the subject was told to try to pick up where he left off in the previous task.

Visual Counting (*Count*): The subject was asked to imagine a blackboard and to visualize numbers being written on the board sequentially, with the previous number being erased before the next number was written and to pick up counting from the previous task.

Hence, the EEG signal of each mental task of 10 sec was divided into nine segments of 1 sec of duration. From the original 10-sec signal, the first and the final 0.5 sec were discarded to avoid the extreme effect of the filtering. The subjects participated in a different number of sessions, i.e. Subject 5 had 135 segments per mental task; Subjects 1 and 3 had 90 segments per mental task; Subjects 2, 6 and 7 had 45 segments per mental task and, finally, Subject 4 had 81 segments per mental task (this last one is due to some errors in the database).

Methods

A scheme of the proposed method is shown in Fig. 1, followed by a description of each block.

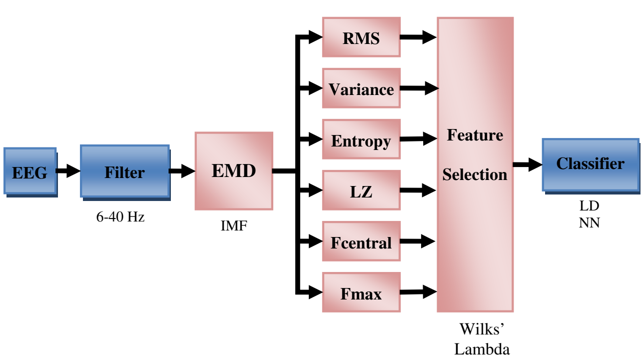

Fig. 1 (Color online) Scheme of the proposed EEG classification method.

Preprocessing

This part of the proposed method is just the filtering stage. The EEG was digitally filtered with a Butterworth bidirectional filter of order 10 with a band-pass between 6 and 40 Hz, aiming at analyzing the α-band, β-band and γ-band. The lower cut-off frequency (6 Hz) was chosen in order to not alter the α-band (8−12 Hz) with the filtering. Thus, a part of the θ-band (< 8 Hz) was included.

Feature extraction

The feature extraction is divided into two parts; the first one is the EMD of the EEG signals, whereas the second part is the estimation of different time and frequency parameters or features.

The empirical mode decomposition

If we assume that any signal is composed of a series of different intrinsic oscillation modes, the EMD[4,5] can be used as a method that carries out this decomposition of the incoming signal into its different intrinsic modes of oscillation. Each oscillatory mode is called an intrinsic mode function (IMF) that can have both the amplitude and the frequency as a function of time. On account of this, the EMD method is very appropriate for nonlinear and non-stationary signals. An IMF is defined as a function that satisfies the following two conditions:

(a) In the entire signal, the number of extremes and the number of zero-crossings must be either equal or differ at most by one, and
(b) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima must be zero (or close to zero).

Given the incoming signal $x(t)$, the algorithm of the EMD is based on a sifting process that can be summarized as follows[4]:

(1) Identify all the local maxima in the signal and then interpolate with a cubic spline line, in order to produce the upper envelope.
(2) Repeat the procedure for the local minima to produce the lower envelope.
(3) Compute the mean of both envelopes that is designated as m_1.
(4) Extract the detail h_1:
$$h_1 = x(t) - m_1. \tag{1}$$
(5) If it is necessary, repeat the Steps 1 to 4, and consider the detail h_i as the data, until detail h_1 can be considered an IMF, satisfying the two above conditions.

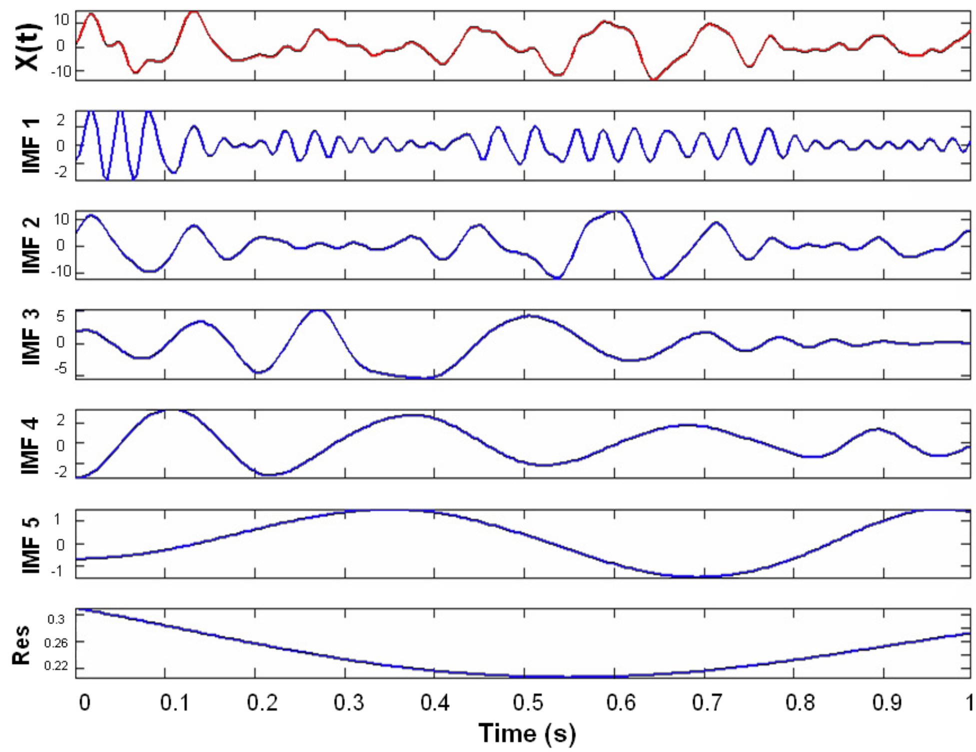

Fig. 2 (Color online) EEG of 1 s decomposed in their respective IMF. $X(t)$ is the incoming EEG signal.

(6) After k iterations, the detail h_k is an IMF and is designated as:

$$\mathrm{IMF}_1 = h_k. \tag{2}$$

(7) Iterate Steps 1 to 6 on the residual r_j in order to obtain all the IMFs of the signal:

$$r_j = x(t) - \mathrm{IMF}_1 - \mathrm{IMF}_2 - \cdots - \mathrm{IMF}_j. \tag{3}$$

The procedure ends when the residual r_j is either a constant, a monotonic slope, or a function with only one extreme.

The result of the EMD process produces n IMFs and a residue signal r_n. The original signal $x(t)$ can be recovered as:

$$x(t) = \sum_{j=1}^{n} \mathrm{IMF}_j + r_n. \tag{4}$$

IMF were extracted using a toolbox for Matlab®, which is available online.[21] Figure 2 illustrates an example of the IMFs of an EEG signal, showing that the lower-order IMFs capture the faster oscillation modes of the signal, whereas the higher-order IMFs capture the slower oscillation modes.

Estimated parameters

For the EEG signals of the database, the EMD algorithm employed usually extracts up to four or five IMFs and the residue (five or six signals in total) per each channel (Fig. 2). Thus, for the analysis the first four IMF and the residue were taken, or the first five IMF, depending on the case. Whether the EMD extracts only three IMFs and the residue, the process creates a fifth IMF with zeros. For each IMF of this five component signals, different parameters can be computed. The proposed parameters utilized in this work are the following:

(1) *Root Mean Square* (*RMS*)
This parameter is calculated as:

$$\mathrm{RMS}_{ij} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} |\mathrm{IMF}_{ij(n)}|^2}, \tag{5}$$

where RMS_{ij} is the RMS of the ith IMF of the jth channel; N is the number of samples of each IMF (250 samples); $\mathrm{IMF}_{ij(n)}$ are the values of the ith IMF of the jth channel at each instant n.

(2) *Variance*
The variance is calculated from each IMF as:

$$\mathrm{Var}_{ij} = \frac{1}{N-1}\sum_{n=1}^{N} (\mathrm{IMF}_{ij(n)} - \overline{\mathrm{IMF}}_{ij(n)})^2, \tag{6}$$

where Var_{ij} is the variance of the ith IMF of the jth channel, $\overline{\mathrm{IMF}}_{ij(n)}$ is the mean value of the IMF.

(3) *Entropy*

Shannon entropy[22] is a measurement of the average amount of information obtained from a measurement. For a system A with N possible states $\{a_1, a_2, \ldots, a_N\}$ each one with its corresponding probability $p(a_i)$, the Shannon entropy of the system H is:

$$H = -\sum_{i=1}^{N} p(a_i) \log p(a_i). \tag{7}$$

For equiprobable events, the entropy is maximal ($H_{\max} = \log(N)$) and, if the probability of an event a_i is one and all the other probabilities are zero, the entropy is minimal ($H_{\min} = 0$). Shannon entropy remains unchanged when adding events with zero probability. Then, the entropy is normalized in terms of $H_{\max}$, according to:

$$H_N = \frac{H}{H_{\max}}. \tag{8}$$

(4) *Lempel–Ziv Complexity Measure*

The Lempel–Ziv (LZ) analysis[23] is a nonlinear method that quantifies the complexity of a time series by analyzing its spatio-temporal patterns. This method is well suited to the analysis of non-stationary biomedical signals of short length and has already been proven as a useful tool to measure the complexity of EEG signals in different areas.[24,25]

LZ complexity analysis is based on a coarse-graining of the analyzed signal, i.e. the signal to be analyzed is ranged into a fixed number φ of quantization levels and is transformed into a sequence of symbols from 1 to φ. The quantization procedure transforms the signal into a sequence of symbols from the limited alphabet of symbols (from 1 to φ). The basic idea of the LZ complexity analysis is to count the number of distinct patterns P contained in the given sequence. The sequence is scanned from left to right and the complexity counter P is increased one unit every time a new subsequence of consecutive characters arises in the scanning process.

Once the number of distinct patterns P has been found, the LZ complexity value of the signal C_φ can be obtained by normalizing P as a function of the length of the analyzed sequence N and the number of quantization levels φ,

$$C_\varphi = \frac{P \log_\varphi(N)}{N}, \tag{9}$$

where $\log_\varphi(N)$ is the φ-base logarithm of N.

(5) *Central Frequency*

Since each IMF is bandwidth limited, it is interesting to measure the frequencies of that band. In that way, the central frequency is proposed as one measurement. The central frequency of each IMF is the frequency that contains half the total spectrum energy, i.e. it splits the spectrum of the IMF into two parts, each one having half the energy. The spectrum is computed in terms of the periodogram with a Hamming window of 512 points.

(6) *Maximum Frequency*

The maximum frequency is the frequency that contains 95% of the energy of the spectrum, and can also be computed with a periodogram with a Hamming window of 512 points.

Feature selection

A disadvantage arising at this point is that the feature vector that would enclose all the features calculated with the above parameters may be too large, i.e. each feature vector contains 180 parameters (5 IMFs $\times$ 6 parameters $\times$ 6 channels). Consequently, a feature selection is essential to do in order to solve this curse-of-dimensionality inconvenience.[26] Therefore, the more important variables for the analysis should be selected, i.e. the variables that contribute with more information. This selection is performed with a *stepwise method* based on the statistical parameter *Wilks' lambda* (WL).[27,28] The WL is chosen to accomplish the feature selection as it is easy and fast to compute. In Appendix A, the WL calculation is detailed.

Classifier

In order to classify the different mental tasks, two different classifiers were implemented; a linear classifier and a nonlinear one with two variations.

Linear Classifier:

A linear discriminant (LD) classifier is the simplest classifier; which consists of a linear combination of variables as stated below:

$$y = \mu_0 + \mu_1 X_1 + \mu_2 X_2 + \cdots + \mu_p X_p, \tag{10}$$

where y is the output value of the discriminant function; μ_i are the coefficients of the discriminant function; X_i are the discriminant variables at each case and p is the number of variables in the analysis.[28]

Nonlinear Classifier:

As a nonlinear classifier, neural networks (NNs) were chosen. A multilayer perceptron[29] with two hidden layers was implemented using Matlab®. Two different configurations were chosen: one with more neurons in the hidden layers than the other one (at each case the number of neurons used is specified in the next section).

The NNs were trained with Levenberg–Marquard backpropagation method, and an early stopping method

Table 1. Accuracy in the Classification of Mental Tasks in *One-Versus-One* Scheme for all Subjects.

	Subject 1			Subject 2			Subject 3			Subject 4			Subject 5			Subject 6			Subject 7		
Mental Tasks Combination	**LD**	**20-10NN**	**10-5NN**	**LD**	**20-10NN**	**10-5NN**	**LD**	**20-10NN**	**10-5NN**	**LD**	**20-10NN**	**10-5NN**	**LD**	**20-10NN**	**10-5NN**	**LD**	**20-10NN**	**10-5NN**	**LD**	**20-10NN**	**10-5NN**
Base-Count	84.86%	88.47%	86.67%	86.95%	79.45%	80.84%	74.03%	70.00%	70.14%	98.44%	92.34%	95.94%	71.67%	70.46%	71.67%	94.86%	82.22%	81.53%	90%	80.83%	79.17%
Base-Lett	95.42%	95.28%	96.11%	85.28%	87.23%	87.23%	78.89%	79.86%	79.17%	97.81%	93.28%	94.69%	84.54%	80.43%	80.19%	95.70%	91.67%	93.33%	99%	95.84%	93.89%
Base-Math	97.64%	95.28%	98.33%	99.17%	82.50%	79.17%	83.61%	82.64%	83.89%	96.56%	92.19%	91.72%	84.16%	87.22%	87.31%	96.25%	92.08%	94.17%	100%	87.23%	84.45%
Base-Rot	98.89%	96.39%	96.95%	89.17%	90.28%	85.28%	74.16%	75.70%	72.50%	86.09%	80.00%	78.59%	83.42%	81.02%	81.76%	97.78%	93.75%	94.17%	92%	91.39%	94.45%
Lett-Count	98.33%	98.33%	98.06%	92.50%	89.45%	92.50%	80.14%	80.14%	76.81%	96.25%	88.91%	91.25%	85.37%	81.29%	81.57%	92.50%	89.03%	86.25%	100%	93.89%	89.72%
Lett-Rot	99.44%	98.19%	98.89%	97.78%	93.06%	93.89%	87.78%	79.03%	81.67%	94.53%	91.25%	93.28%	95.47%	90.93%	90.37%	96.11%	91.39%	92.08%	100%	95.84%	94.72%
Math-Count	98.89%	95.83%	96.94%	88.61%	85.00%	83.61%	82.50%	78.75%	79.03%	90.94%	86.72%	90.16%	82.13%	83.43%	82.78%	90.42%	88.47%	90.70%	99.17%	83.89%	81.94%
Math-Lett	100%	97.92%	97.64%	98.89%	84.72%	82.23%	87.64%	83.89%	86.11%	95%	92.97%	90.31%	85.28%	80.09%	81.02%	94.72%	84.86%	90.00%	99%	97.50%	96.39%
Math-Rot	100%	95.56%	96.95%	96.12%	94.45%	92.50%	87.36%	82.50%	81.11%	91%	90.16%	92.34%	92.96%	91.48%	93.98%	94.30%	88.89%	91.25%	100%	93.06%	89.45%
Rot-Count	76.67%	73.06%	70.83%	90.00%	90.28%	88.06%	73.20%	71.53%	70.70%	80.16%	81.25%	82.03%	86.30%	81.94%	82.59%	93.47%	86.95%	87.08%	94%	88.06%	92.50%
Average	**95.00%**	**93.43%**	**93.74%**	**92.45%**	**87.64%**	**86.53%**	**80.93%**	**78.40%**	**78.11%**	**92.73%**	**92.73%**	**90.03%**	**85.13%**	**82.83%**	**83.32%**	**94.61%**	**88.93%**	**90.06%**	**97.36%**	**90.75%**	**89.67%**
ITR [bits/trial]	**0.71**	**0.65**	**0.66**	**0.61**	**0.46**	**0.43**	**0.3**	**0.25**	**0.24**	**0.624**	**0.624**	**0.53**	**0.39**	**0.34**	**0.35**	**0.7**	**0.5**	**0.53**	**0.82**	**0.56**	**0.52**

was used to stop the training process. In the output layer, one neuron per each mental state was utilized.

RESULTS

Two approaches were proposed to classify the mental tasks, in the first approach the *one-versus-one* classification scheme (mental tasks classified in pairs) was used and in the second one an *all-versus-all* scheme was applied, i.e. the five mental tasks at the same time.

The feature selection is an important issue in order to solve the curse of dimensionality[27] and with the application of WL value, the initial feature vector (containing 180 parameters) is reduced to a small number of only 16 ± 7 parameters (depending on the subject and the mental tasks) in the *one-versus-one* classification scheme. In the *all-versus-all* scheme, the feature vector was reduced to 15 ± 5 variables. From the whole variables chosen in the analysis in the *all-versus-all* scheme, the RMS parameters (from different IMFs and EEG channels) represented the 31% of the chosen variables. On the other hand, variance, LZ, entropy, maximum and central frequencies parameters were respectively chosen in the 18%, 16%, 12%, 12% and 11% of the cases. Generally, only the foremost IMFs (first to third) were chosen in the analysis.

Table 1 shows the results obtained in the *one-versus-one* scheme for each subject. These values are obtained using a 10-fold cross-validation. In order to obtain more accurate results, the cross-validation method was repeated over four times.

As a result, the values shown in this table are the average values over results obtained in each cross-validation. In this table, the NNs had the same number of neurons as the number of parameters selected by the WL method in its input, two neurons in its output, and the two chosen configurations were the 10-5NN (10 and 5 neurons in hidden layers) and 20-10NN (20 and 10 neurons in hidden layers). Furthermore, in Table 1 the average values are expressed as bit rate (BR) in bits/trial. The BR expresses the human−machine channel capacity to transfer information. It is calculated as[30]:

$$\mathrm{BR} = \log_2(N) + P\log_2(P) + (1-P)\log_2\left(\frac{1-P}{N-1}\right). \quad (11)$$

The average classification over all subjects with LD was 91.17% (SD: 5.3%), with 20-10NN was 87.82% (SD: 5.3%) and with 10-5NN was 87.35% (SD: 5.9%). Additionally, the average values of each subject and the mean over all subjects are presented in Fig. 3.

In Fig. 4, the results of *all-versus-all* scheme are expressed as confusion matrices, but only the results obtained with the best classifier per subject are presented. Once again, three classifiers were used: a LD and the NN with two different configurations (one with 20 and 10 neurons and the other with 40 and 20 neurons in their hidden layers, respectively), both NNs with 5 neurons in their output layer. The results are obtained with 10-fold cross-validation repeated over four times.

In a multi-class classification problem, the proper evaluation of the classifier is described by its confusion matrix, which is difficult to evaluate and to compare results. Generally, the accuracy classification (ACC), calculated from the diagonal values of the confusion matrix, is reported. Table 2 presents the overall ACC of the three classifiers for each subject. However, the ACC does not describe the entire result since the values outputted from the diagonal are not used. Therefore,

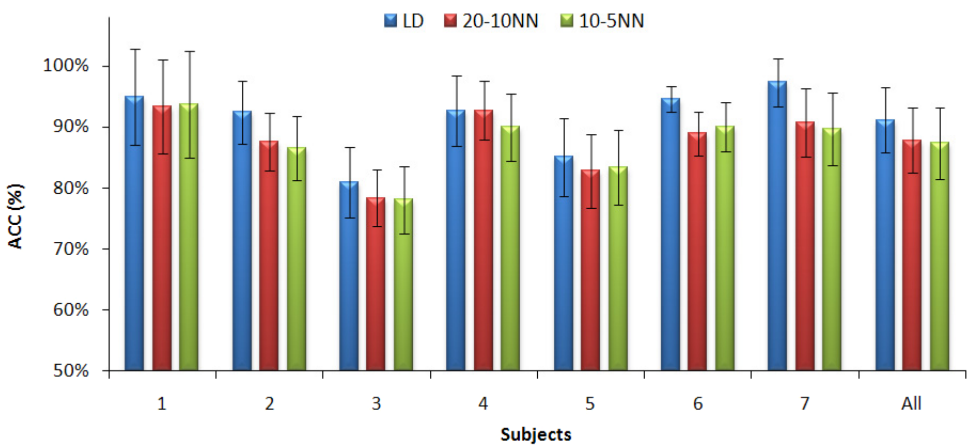

Fig. 3 (Color online) Overall results of each subject and for all subjects in the *one-versus-one* classification scheme. The values are expressed as average results and error bars represent the standard deviation. LD: 20-10NN: neural network with 20 and 10 neurons per hidden layers; 10-5NN: neural network with 10 and 5 neurons per hidden layers.

Subject 1 (40-20NN)

	Base	Count	Lett	Mult	Rot
Base	**88.06**	9.44	0.56	0.28	1.67
Count	16.94	**42.22**	0.28	1.67	38.89
Lett	0.56	0.56	**97.5**	0	1.39
Mult	1.39	1.11	0	**97.5**	0
Rot	2.78	29.44	0.83	0.56	**66.39**

Subject 2 (LD)

	Base	Count	Lett	Mult	Rot
Base	**47.16**	28.98	6.25	5.11	13.64
Count	17.61	**73.86**	3.41	0	6.25
Lett	16.48	15.91	**65.34**	0.57	0.00
Mult	10.80	11.36	2.27	**73.3**	0
Rot	8.52	13.64	0	0	**79.55**

Subject 3 (LD)

	Base	Count	Lett	Mult	Rot
Base	**47.50**	25.28	3.89	7.5	15.83
Count	16.67	**46.11**	9.17	6.94	21.11
Lett	16.39	8.33	**54.17**	8.89	12.22
Mult	15.83	8.89	7.5	**56.11**	11.67
Rot	19.72	17.78	4.44	10.83	**47.22**

Subject 4 (20-10NN)

	Base	Count	Lett	Mult	Rot
Base	**73.75**	4.06	0.94	3.13	18.44
Count	6.88	**67.50**	2.81	5.31	16.56
Lett	2.81	5.31	**85.31**	4.06	3.13
Mult	3.13	4.38	4.38	**84.06**	4.06
Rot	22.19	21.88	3.13	2.5	**50.31**

Subject 5 (20-10NN)

	Base	Count	Lett	Mult	Rot
Base	**47.20**	22.2	7.65	6.34	16.98
Count	22.39	**52.80**	7.84	5.6	11.38
Lett	5.6	4.29	**78.73**	9.7	2.24
Mult	7.65	6.34	14.18	**67.72**	3.73
Rot	9.52	12.13	2.24	2.99	**72.57**

Subject 6 (LD)

	Base	Count	Lett	Mult	Rot
Base	**70.83**	12.22	10	4.17	2.78
Count	12.22	**67.78**	1.11	2.5	16.39
Lett	6.67	5.56	**76.67**	5.56	5.56
Mult	0.28	7.50	6.67	**80.83**	4.72
Rot	5.56	6.67	0	2.78	**85**

Subject 7 (LD)

	Base	Count	Lett	Mult	Rot
Base	**63.07**	28.41	0	0.57	7.39
Count	27.84	**67.05**	0	0	6.82
Lett	2.84	5.11	**90.34**	0	2.84
Mult	6.82	2.27	0	**87.50**	1.70
Rot	2.27	6.82	0	0	**90.34**

Fig. 4 (Color online) Confusion matrices of each subject. Only the best results obtained for each subject are shown. The parenthesis indicates the classifier used to obtain these results.

Table 2. Overall Results and Kappa Coefficient for Each Subject in the *All-Versus-All* Scheme.

	LD		**40-20NN**		**20-10NN**	
Subjects	**ACC**	$\kappa \pm \mathrm{Se}(\kappa)$	**ACC**	$\kappa \pm \mathrm{Se}(\kappa)$	**ACC**	$\kappa \pm \mathrm{Se}(\kappa)$
1	72.83%	0.66 ± 0.048	78.33%	0.729 ± 0.051	78.28%	0.728 ± 0.051
2	67.84%	0.598 ± 0.066	58.64%	0.483 ± 0.061	59.32%	0.491 ± 0.062
3	50.22%	0.378 ± 0.04	45.06%	0.313 ± 0.038	45.50%	0.319 ± 0.038
4	71.94%	0.649 ± 0.051	68.38%	0.605 ± 0.05	72.19%	0.652 ± 0.051
5	59.74%	0.497 ± 0.036	57.01%	0.463 ± 0.035	63.81%	0.548 ± 0.037
6	76.22%	0.703 ± 0.05	69.83%	0.623 ± 0.048	73.33%	0.667 ± 0.049
7	79.66%	0.746 ± 0.073	70.11%	0.626 ± 0.068	70.45%	0.631 ± 0.068
Average	**68.35%**	**0.634 ± 0.054**	**63.91%**	**0.571 ± 0.052**	**66.13%**	**0.593 ± 0.053**

Table 3. Accuracy Classification and Cohen's Kappa for Some Combination of the Extracted Features per Subject in an *All-Versus-All* Scheme.

	RMS		Var		Ent		Fc		Fmax		LZ		RMS-Ent	
Subject	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**
1	49.78%	0.37	49.78%	0.37	52.22%	0.40	23.78%	0.05	26.89%	0.09	53.78%	0.42	73.8%	0.67
2	40.44%	0.26	42.67%	0.28	43.56%	0.29	33.78%	0.17	32.89%	0.16	39.11%	0.24	59.1%	0.49
3	35.78%	0.20	35.78%	0.20	36.00%	0.20	31.78%	0.15	30.00%	0.13	38.44%	0.23	47.8%	0.35
4	51.61%	0.40	51.61%	0.40	45.43%	0.32	37.28%	0.22	32.84%	0.16	47.16%	0.34	66.2%	0.58
5	44.74%	0.31	44.00%	0.30	44.30%	0.30	28.30%	0.10	25.93%	0.07	44.44%	0.31	56.4%	0.46
6	56.00%	0.45	52.89%	0.41	58.67%	0.48	43.11%	0.29	43.11%	0.29	60.00%	0.50	73.1%	0.66
7	46.22%	0.33	44.00%	0.30	42.22%	0.28	36.89%	0.21	40.89%	0.26	51.56%	0.39	67.1%	0.59
Average	**46.37%**	**0.33**	**45.82%**	**0.32**	**46.06%**	**0.33**	**33.56%**	**0.17**	**33.22%**	**0.17**	**47.78%**	**0.35**	**63.4%**	**0.54**

	RMS-Fmax		Var-Ent-LZ		RMS-LZ		RMS-Ent-Fc-Fmax-LZ		Var-Ent-Fc-Fmax-LZ		RMS-Var-Ent-Fmax-LZ		RMS-Var-Ent-Fc-Fmax-LZ	
	ACC	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**	**ACC**	**κ**
1	53.8%	0.42	51.8%	0.40	72.22%	0.653	74.4%	0.68	53.8%	0.42	72.9%	0.66	72.9%	0.66
2	46.7%	0.33	43.6%	0.29	60.44%	0.506	64.9%	0.56	57.8%	0.47	65.8%	0.57	**64.9%**	0.56
3	36.7%	0.21	35.8%	0.20	46.67%	0.333	47.8%	0.35	46.0%	0.33	47.1%	0.34	**52.9%**	0.41
4	53.3%	0.42	52.6%	0.41	65.93%	0.574	**70.4%**	0.63	60.7%	0.51	63.5%	0.54	69.9%	0.62
5	46.4%	0.33	47.4%	0.34	55.56%	0.444	59.7%	0.50	49.8%	0.37	54.7%	0.43	**60.1%**	0.50
6	61.3%	0.52	63.8%	0.55	71.56%	0.644	**76.4%**	0.71	72.2%	0.65	75.8%	0.70	75.3%	0.69
7	50.7%	0.38	51.6%	0.39	70.22%	0.628	75.6%	0.69	69.3%	0.62	**77.3%**	0.72	76.4%	0.71
Average	**49.8%**	**0.37**	**49.5%**	**0.37**	**63.23%**	**0.54**	**67.0%**	**0.59**	**58.5%**	**0.48**	**65.3%**	**0.57**	**67.5%**	**0.59**

another measurement for precision is used in Table 2, namely, the *Cohen's kappa coefficient*.[31] This coefficient has already been proven established in other areas of EEG research, such as in sleep classification research.[32] In the BCI area, the kappa coefficient was introduced in Ref. 33. The kappa coefficient takes values between 0 and 1 and can be interpreted on the basis of its value. Therefore, values ranging from 0 to 0.20 mean a slight agreement; from 0.21 to 0.40 represent a fair agreement; from 0.41 to 0.60 indicate a moderate agreement; from 0.61 to 0.80 imply a substantial agreement and values between 0.81 and 1.00 represent an almost perfect agreement. Although, this scale is not universally accepted, it renders a panorama on agreement levels in the classification. More details about Cohen's kappa coefficient computation are presented in Appendix B at the end of this paper.

Another issue analyzed in this work is the amount of information given by each calculated parameter (RMS, variance, Shannon entropy, LZ, central frequency and maximum frequency). For this reason, these parameters were tested independently, and combinations of them were tested as well. Table 3 shows these results obtained with the LD in *all-versus-all* scheme. Table 3 presents the most representative combinations (14 out of 63). The remainder combinations are not shown due to their poor results achieved and space limitations. The feature selection using WL parameter is applied in these cases as well, but the feature vector has 30 components for the cases of independent parameters, and more components for the combination among themselves (60 components for two parameters, 90 components for three parameters, etc.). The reduction of components varies for each subject and for each parameter.

DISCUSSION

The *one-versus-one* classification scheme (shown in Table 1) indicates that the Subject 1 obtained, for all the classifiers, results above 80% for all the possible combinations, excepting for the Rot-Count combination with lower results, though greater than 70%. Indeed, for some combinations, a value of 100% of ACC is attained. For Subject 2, a similar behavior is observed; i.e. greater results to 83% are obtained using any classifier, excepting for the NNs which obtained, for some combinations, lower results (but greater than 79%). Subject 3 showed the worst overall performance, with results lying between 70% and 87%, the last one was obtained with the LD. In the case of Subject 4, the results are similar to those of Subjects 1 and 2. Subject 5 obtained a low result in Base-Count combination (70%), but in the remaining combinations, it attained values greater than 80%. Subjects 6 and 7 had an average performance of 94% and 97% with LD, respectively; and 89% with the NNs as a lower performance. The average BR was ranging from 0.24 bits/trials up to 0.84 bits/trial.

In the majority of mental tasks combinations, the best results were obtained using LD for all subjects. This fact is easy to note in Fig. 3, where the averaged ACC results of each subjects are presented. An overall average over all subjects shows very high performances: 91% for LD and 87% for both NNs. The ACC attained with this method were higher than the results documented in similar works with the same database.[34–36] In Ref. 34, power and asymmetry ratios of EEG bands were used with an average ACC of 86.5% on four subjects with an Elman NN. In Ref. 35, autoregressive (AR) modeling and multilayer perceptron were used with ACC up to 71%. The best result obtained with AR modeling and support vector machines was 72%.[36] Recently, the bivariate extension of EMD (BEMD) was used to obtain a more accurate estimate of the marginalized spectrum for the calculation of amplitude asymmetry in frequency.[17] In that work, the proposed method was tested on four subjects extracted from the same database used in the current paper; an average ACC of mental task of 69.05% was achieved.

In the *all-versus-all* classification scheme illustrated in Fig. 4 and presented in Table 2, the best ACC is attained with LD for Subjects 2, 3, 6 and 7, with NN-20-10 for Subjects 4 and 5, and only for Subject 1 the NN-40-20 obtained the best performance. In this scheme, the superiority of the LD is not very obvious over the NNs. Again, Subject 3 obtained the worst results. In the confusion matrices, we can see that the Count mental task is greatly confused with the Base and Rot mental tasks. Indeed, the greatest detriment in the results comes from this fact. Subjects 1, 3, 5 and 6 confuse Count mainly with the other two mental tasks, whereas Subjects 2 and 7 confuse Count with Base and Subject 4 confuses with Rot. The acquisition of EEG signals with more electrodes located in other positions on the scalp probably improves the ACC of the confused mental tasks (Count, Base and Rot).

From Table 2, one can perceive that the kappa coefficients resulting from the confusion matrices show that Subjects 1, 4, 6 and 7 present a substantial agreement ($0.61 < \kappa < 0.81$) in the classification. This is in accordance to ACC values from the first column of Table 2, since those subject obtained ACC higher than 70%. Subject 2, with κ a bit smaller than 0.61, can be considered an acceptable result ($\text{ACC} = 67.84\%$). Whereas Subject 5 obtained $\text{ACC} = 59.74\%$ with $\kappa = 0.49$ indicating a moderate agreement. Finally, Subject 3 did not perform that well ($\text{ACC} = 50.22\%$ with $0.21 < \kappa < 0.40$ representing a fair agreement).

Table 3 presents an analysis of the amount of information added for each parameter and for several combinations of them. The parameters RMS, variance, entropy and LZ had a similar performance, and the frequencial parameters obtain lower results. The combination of two or more parameters brings along better results than individual parameters by themselves. Another fact, worth noting is that when the frequency parameters are added to any combination of the other parameters, the results were improved. The variation of frequency parameters was analyzed in a previous work.[35] The best ACC was obtained using the combination of all parameters, although, using only a few parameters, it is possible to obtain very similar results. For instance, the combinations of *RMS-Ent-Fc-Fmax-LZ* and *RMS-Var-Ent-Fc-LZ* allow obtaining average accuracy of 67% with $\kappa = 0.59$ in both cases. Moreover, *RMS-Var-Ent-Fmax-LZ* obtained 65.3% ($\kappa = 0.57$) in average, whereas the combination of *RMS-Var-Ent-fc-fmax* achieved 66.7% ($\kappa = 0.58$), as well.

The EMD has the problem of mode mixing and it is improved using the ensemble EMD (EEMD), which uses ensemble of mean of IMFs decomposed from noise-added signals. However, EEMD is more time demanding than EMD thus EEMD could not be suitable for real-time BCI applications.

In this work, it has been found that the WL parameter allows choosing the more suitable variables in the analysis, and allows solving the curse-of-dimensionality by reducing the feature vector of 180 variables into a small number of 16 ± 7 variables for each combination in the *one-versus-one* scheme, which allows a better ACC. Generally, only the foremost IMFs (first to third) were selected by WL, i.e. it is not necessary to extract all the IMFs of the signal. The use of the feature selection in order to solve the curse-of-dimensionality problem of the feature vector is an important aspect of BCI applications.[27] The feature selection through the WL parameter accomplishes this objective, while being easy and fast to compute as well. In the *all-versus-all* scheme, the RMS represents the most frequently chosen parameter (31% of times), followed by the variance and the LZ, with 18% and 16% of times, respectively. The other parameters were chosen 11% or 12% of times. That means that it is not necessary to compute all the proposed parameters because, by using some combination of parameters, it is possible to reach similar results in the ACC (see Table 3). Hence, any combination of parameters should contain the RMS value since it is the most important parameter in this analysis. The other parameters collaborate in obtaining higher results in the classification and thus, some combinations could achieve similar results.

CONCLUSIONS

In this work, a features extraction method is proposed for the processing of EEG signals and classification of mental tasks. It is based on the EMD and the estimation of several parameters, namely RMS, variance, Shannon entropy, LZ complexity value, and central and maximum frequencies. A reduction of dimensionality was performed, based on the WL parameter. Two classification schemes of mental tasks were tested, the *one-versus-one* and the *all-versus-all* scheme. For both schemes, two different classifiers (LD and NN) were employed. In the last one, different numbers of neurons per layer were tested.

This method allows attaining very high results in the ACC of mental tasks, obtaining performances greater than 90% for almost all subjects in the *one-versus-one* scheme using any classifier. In the *all-versus-all* scheme, the method reached an ACC higher than 70% for almost all cases. The LD performs better than NN with this method, and there was no difference between the different topologies of NN used. The results were higher than those documented in similar works using the same database.

In future, the proposed method could be implemented for the analysis of the ongoing EEG aimed to controlling a BCI.

ACKNOWLEDGMENTS

The authors want to thank C.W. Anderson for providing the EEG database used in this work and to P. Flandrin *et al.* for the EMD toolbox. The first three authors are supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

APPENDIX A. WILKS' LAMBDA

WL measures the ratio of within-group variability respecting the total variability on the discriminator variables, and it is a measurement of the importance of the functions. In this inverse measurement scale, values close to 1 indicate that almost all of the variability in the discriminator variables originates from within-group differences (i.e. differences between cases in each group), whereas values close to 0 indicate that almost all of the variability in the discriminator variables is due to group differences.[27]

The WL, in a p-dimensional space constructed with p variables and with the matrices B_{pxp} and W_{pxp} can be defined as the ratio between their determinants[28]:

$$WL = \frac{|W|}{|B+W|}, \tag{A.1}$$

where B and W are matrices whose elements represent the square sum and cross products within-groups and between groups, respectively. Then, the value of WL is transformed into the general multivariate statistical F, which allows contrasting significant differences between groups:

$$F = \frac{n-g-s}{g-1}\left(\frac{1-\Lambda_{s+1}/\Lambda_s}{\Lambda_{s+1}/\Lambda_s}\right), \tag{A.2}$$

where, n, g and s are the number of cases, groups and selected variables respectively; Λ_s is the WL before adding a new variable, and Λ_{s+1} results after adding that variable. To accept a variable in the analysis, the *F value* must be higher than 3.84 (namely, "F to enter") and, once included, the variable is rejected if its *F value* is smaller than 2.71 (namely, "F to exit").

APPENDIX B. KAPPA COEFFICIENT

The kappa coefficient[31] is defined, for an M-class classification problem, from the confusion matrix M. Hence, we can derive the ACC (overall agreement) as:

$$\mathrm{ACC} = p_0 = \frac{\sum_i M_{ii}}{N} \tag{B.1}$$

and the chance expected agreement is:

$$p_e = \frac{\sum_i n_{oi} \times n_{io}}{N^2}, \tag{B.2}$$

where N is the total number of samples, M_{ii} are the elements on the main diagonal of M, n_{oi} and n_{io} are the sums of each column and each row, respectively. Then, the estimate of the kappa coefficient κ is:

$$\kappa = \frac{p_0 - p_e}{1-p_e} \tag{B.3}$$

and its standard error $\mathrm{Se}(\kappa)$ is measured as:

$$\mathrm{Se}(\kappa) = \frac{\sqrt{p_0 + p_e^2 - \sum_i [n_{oi} \times n_{io} \times (n_{oi} + n_{io})]/N^3}}{(1-p_e)\sqrt{N}}. \tag{B.4}$$

REFERENCES

1. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM, Brain-computer interfaces for communication and control, *Electroen Clin Neuro* **113**:767−791, 2002.

2. Sörnmo L, Laguna P, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, 1st ed., Elsevier Inc., USA, 2005.
3. McFarland DJ, Anderson CW, Müller KR, Schlögl A, Krusienski DJ, BCI meeting 2005 — Workshop on BCI signal processing: Feature extraction and translation, *IEEE Trans Neural Syst Rehabil Eng* **14**:135−138, 2006.
4. Huang NE, Shen Z, Long SR, Wu ML, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH, The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis, *Proc R Soc London A* **454**:903−995, 1998.
5. Huang NE, Shen SSP (eds.), Hilbert-Huang transform and its applications, *Interdisciplinary Mathematical Sciences*, World Scientific Publishing, pp. 1−26, 2005.
6. Jing-tian T, Qing T, Yan T, Bin L, Xiao-kai Z, Hilbert-Huang Transform for ECG De-noising, *Proc 1st Int Conf on Bioinformatics and Biomed Eng (ICBBE 2007)*, pp. 664−667, 2007.
7. Liang H, Bressler SL, Desimone R, Fries P, Empirical mode decomposition: A method for analyzing neural data, *Neurocomputing* **65**:801−807, 2005.
8. Torres A, Jané R, Fiz JA, Laciar E, Galdiz JB, Gea J, Morera J, Analysis of respiratory mechanomyographic signals by means of the empirical mode decomposition, *J Phys: Conf Ser* **90**:1−8, 2007.
9. Hu M, Li G, Ding Q, Li J, Classification of normal and hypoxia EEG based on Hilbert-Huang transform, *Proc Int Conf on Neural Net and Brain (ICNN&B'05)*, Vol. 2, pp. 851−854, 2005.
10. Orosco L, Garcés Correa A, Laciar E, Multiparametric detection of epileptic seizures using empirical mode decomposition of EEG records, *Proc Annual Int Conf of the IEEE EMBS (EMBC 2010)*, pp. 951−954, 2010.
11. Rutkowsk TM, Zdunek R, Cichocki A, Multichannel EEG brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support, *Proc Int Congr Ser* **1301**:266−269, 2007.
12. Solis-Escalante T, Gentiletti GG, Yañez-Suarez O, Single Trial P300 detection based on the Empirical Mode Decomposition, *Proc 28th IEEE EMBS Annual Int Conf*, pp. 1157−1160, 2006.
13. Wan B, Zhou Z, Xu L, Ming D, Qi H, Cheng L, Mu Rhythm Desynchronization Detection Based on Empirical Mode Decomposition, *Proc 31st Annual Int Conf of the IEEE EMBS*, pp. 2232−2235, 2009.
14. Yeh C-L, Chang H-C, Wu C-H, Lee P-L, Extraction of single-trial cortical beta oscillatory activities in EEG signals using empirical mode decomposition, *BioMed Eng OnLine* **9**:1−17, 2010.
15. Trad D, Al-ani T, Monacelli E, Jemni M, Nonlinear and nonstationary framework for feature extraction and classification of motor imagery, *Proc IEEE Int Conf on Rehab Robotics*, 2011.
16. Yuan L, Yang B, Ma S, Cen B, Combination of wavelet packet transform and Hilbert-Huang transform for recognition of continuous EEG in BCIs, *Proc 2nd IEEE Int Conf on Computer Science and Information Technology (ICCSIT 2009)*, pp. 594−599, 2009.
17. Park C, Looney D, Kidmose P, Ungstrup M, Mandic DP, Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition, *IEEE Trans on Neural Syst and Rehab Eng* **19**:366−373, 2011.
18. Diez PF, Mut V, Laciar E, Torres A, Avila E, Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification, *Proc 31st Annual Int Conf of the IEEE EMBS*, pp. 2579−2582, 2009.
19. Keirn ZA, Aunon JI, A new mode of communication between man and his surroundings, *IEEE Trans Biomed Eng* **37**:1209−1214, 1990.
20. Anderson CW, EEG Pattern Analysis. Last accessed March 5, 2012. Available: http://www.cs.colostate.edu/eeg/index.html#Data.
21. Rilling G, EMD Matlab 7.1 codes with examples. Last accessed March 5, 2012. Available: http://perso.ens-lyon.fr/patrick.flandrin/emd.html.
22. Shannon CE, A mathematical theory of communication, *AT&T Tech J* **27**:379−423, 623−656, 1948.
23. Lempel A, Ziv J, On the complexity of finite sequences, *IEEE Trans Inform Theory* **IT-22**:75−81, 1976.
24. Zhang XS, Roy RJ, Jensen EW, EEG complexity as a measure of depth anesthesia for patients, *IEEE Trans Biomed Eng* **48**:1424−1433, 2001.
25. James CJ, Abásolo D, Gupta D, Space-time ICA versus Ensemble ICA for ictal EEG analysis with component differentiation via Lempel-Ziv complexity, *Proc 29th IEEE EMBS Annual International Conference*, pp. 5473−5476, 2007.
26. Lotte F, Congedo M, L'ecuyer A, Lamarche F, Arnaldi B, A review of classification algorithms for EEG-based brain−computer interfaces, *J Neural Eng* **4**:R1−R13, 2007.
27. Tinsley HEA, Brown SD (eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, 1st edn., Academic Press, USA, 2000.
28. Gil Flores J, García Giménez E, Rodríguez Gomez G, *Books of Statistics N°12: Discriminant Analysis* (in Spanish). Ed. La Muralla S.A. and Ed. Hespérides S.I. 2001.
29. Haykin S, *Neural Networks A Comprehensive Foundation*, Pretince Hall, New Jersey, USA, 1999.
30. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM, Brain−computer interface technology: A review of the first international meeting, *IEEE Trans Rehab Eng* **8**(2):164−173, 2000.
31. Cohen J, A coefficient of agreement for nominal scales, *Educ Psychol Meas* **20**:37−46, 1960.
32. Anderer P, Gruber G, Parapatics S, Woertz M, Miazhynskaia T, Klosch G, Saletu B, Zeitlhofer J, Barbanoj MJ, Danker-Hopfe H, Himanen SL, Kemp B, Penzel T, Grozinger M, Kunz D, Rappelsberger P, Schlogl A, Dorffner G, An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: Validation study of the Somnolyzer 24 × 7 utilizing the siesta database, *Neuropsychobiology* **51**:115−133, 2005.
33. Schlögl A, Lee F, Bischof H, Pfurtscheller G, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, *J Neural Eng* **2**:L14−L22, 2005.

34. Palaniappan R, Utilizing gamma band to improve mental task based brain-computer interface design, *IEEE Trans Neural Syst Rehabil Eng* **14**(3):299–303, 2006.
35. Anderson CW, Sijercic Z, Classification of EEG signals from four subjects during five mental tasks, *Proc Int Conf on Eng Appl of Neural Net: Solving Eng Problems with Neural Networks (EANN'96)*, 1996.
36. Garrett D, Peterson DA, Anderson CW, Thaut MH, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, *IEEE Trans Neural Syst Rehab Eng* **11**:141–144, 2003.
37. Diez PF, Laciar E, Torres A, Mut V, Avila E, Adaptive bands on EEG signals extracted with empirical mode decomposition, *V Latin American Congress on Biomedical Engineering CLAIB 2011*, IFMBE Proceedings 33.