31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Application of the Empirical Mode Decomposition to the Extraction of Features from EEG Signals for Mental Task Classification

Pablo F. Diez, Vicente Mut, Eric Laciar, *Member IEEE,* Abel Torres, *Member IEEE,* Enrique Avila

Abstract— **In this work, it is proposed a technique for the feature extraction of electroencephalographic (EEG) signals for classification of mental tasks which is an important part in the development of Brain Computer Interfaces (BCI). The Empirical Mode Decomposition (EMD) is a method capable to process nonstationary and nonlinear signals as the EEG. This technique was applied in EEG signals of 7 subjects performing 5 mental tasks. For each mode obtained from the EMD and each EEG channel were computed six features: Root Mean Square (RMS), Variance, Shannon Entropy, Lempel-Ziv Complexity Value, and Central and Maximum Frequencies, obtaining a feature vector of 180 components. The *Wilks' lambda* parameter was applied for the selection of the most important variables reducing the dimensionality of the feature vector. The classification of mental tasks was performed using Linear Discriminate Analysis (LD) and Neural Networks (NN). With this method, the average classification over all subjects in database was 91±5% and 87±5% using LD and NN, respectively. It was concluded that the EMD allows getting better performances in the classification of mental tasks than the obtained with other traditional methods, like spectral analysis.**

I. Introduction

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]. The feature extraction is an important part in BCI development. In this work is proposed a method to extract features from electroencephalographic EEG signals based on the Empirical Mode Decomposition (EMD). A review of signal processing techniques in feature extraction on EEG signal for BCI is shown in [2]. The EMD is an innovative technique for the analysis of nonlinear and nonstationary time series [3], such as EEG. Hence, EMD is proposed as a new tool for feature extraction of EEG signals for BCI applications.

The results achieved with EMD are higher than those reported in other works with traditional EEG signal processing, utilizing the same database [4-7].

II. EEG Database

The EEG database utilized in this work was acquired by Keirn and Aunon (Colorado State University) [4] and is available on-line [8]. Electrodes were placed at O_1, O_2, P_3, P_4, C_3 and C_4 and referenced to A_1 and A_2. Bandpass analog filters were set at 0.1–100 Hz. Signals were recorded for 10 s during each task and each task was repeated for ten sessions. Seven subjects, 21 to 48 years old, participated in the study involving a total of five distinct tasks, namely:

Baseline Task (Base): The subject was told to simply relax and try to think of nothing in particular.

Mathematical Multiplication Task (Math): The subject was given a nontrivial multiplication problem to solve.

Geometric Figure Rotation (Rot): The subject had to visualize a 3D object being rotated about an axis.

Mental Letter Composing (Lett): The subject was instructed to mentally compose a letter to a friend or relative.

Visual Counting (Count): The subject was asked to imagine a blackboard with numbers being written on it.

The subjects were instructed to not vocalize or make overt movements while solving the tasks.

Hence, the EEG signal of each mental task of 10 s was divided into 9 segments of 1 s of duration. The first and the final 0.5 s were discarded to avoid the extreme effect of the filtering. The subjects participated in a different number of sessions, resulting in different number of segments per mental task, i.e. Subject 5 (135 segments); Subjects 1, 3 and 6 (90 segments); Subjects 2 and 7 (45 segments) and Subject 4 (81 segments).

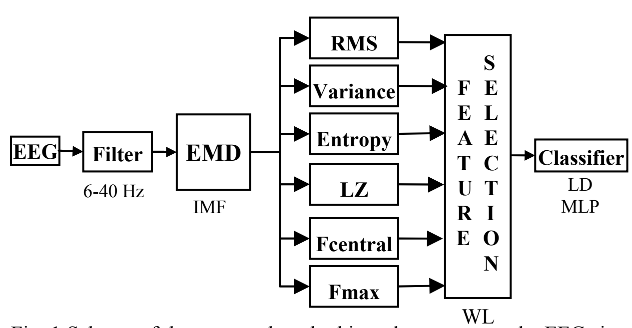

Fig. 1 Scheme of the proposed method in order to process the EEG signals on the proposed EEG classification method

This work was supported in part by grants of Ministerio de Ciencia e Innovación de España (TEC2007-68076-C02-01). The first, second and third author are supported by CONICET of Argentina.

P. F. Diez and E. Laciar are with Gabinete de Tecnología Médica (GATEME), Universidad Nacional de San Juan (UNSJ), Argentina. (e-mail: pdiez@gateme.unsj.edu.ar, laciar@gateme.unsj.edu.ar).

V. Mut is with Instituto de Automática, Facultad de Ingeniería. UNSJ, San Juan, Argentina. (e-mail: vmut@inaut.unsj.edu.ar).

P. F. Diez and E. Avila are with Laboratorio de Electrónica Digital (LED), Facultad de Ingeniería. UNSJ, Argentina (e-mail: eavila@inaut.unsj.edu.ar)

A. Torres is with Dept. ESAII, Universitat Politècnica de Catalunya, Institut de Bioenginyeria de Catalunya (IBEC) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN),Barcelona, Spain (e-mail: abel.torres@upc.edu).

III. METHODOLOGY

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

A. Preprocessing

EEG signals were digitally filtered with a Butterworth bidirectional filter of order 5 with a passband between 6 Hz and 40 Hz, in order to analyze the α, β and γ bands.

B. Feature extraction

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

1) The Empirical Mode Decomposition

If we assume that any signal is composed of a series of different intrinsic oscillation modes, the EMD [3] can be used as a method that carries out this decomposition of the incoming signal into its different Intrinsic Mode Function (IMF). An IMF is a function that satisfies two conditions:

1. In the entire signal, the number of extremes and the zero-crossings must be equal or differ at most by one.
2. 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 it).

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

1. Interpolate all the local maxima and minima in the signal with a cubic spline line, to produce the upper and lower envelope.
2. Repeat for the local minima to produce the lower envelope.
3. Compute the mean of both envelopes m_1.
4. Extract the detail $h_1 = x(t) - m_1$ (1)
5. Repeat the steps 1 to 4, and consider the detail h_i as the data, until detail h_1 can be considered an IMF.
6. After k iterations, the detail h_k is an IMF and is designated as: $IMF_1 = h_k$ (2)
7. Iterate steps 1 to 6 on the residual r_j in order to obtain all the IMFs of the signal:

$$r_j = x(t) - IMF_1 - IMF_2 - \ldots - IMF_j \quad (3)$$

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 summing up the n extracted IMF and the residue:

$$x(t) = \sum_{j=1}^{n} IMF_j + r_n \quad (4)$$

In order to obtain the IMFs of the signal an EMD toolbox for Matlab® was utilized and is available on line [9]. Fig. 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. As it can be seen, the IMF 4 and 5 show frequencial components that supposedly would have been eliminated in the filtering stage, but the amplitude of these IMFs are 20 times minor than IMF 1. This effect may be attributed to some residual components on the filtering stage. Besides, generally, these IMFs are discarded in the feature selection process.

2) Estimated Parameters

The EMD algorithm was applied to each EEG 1 s segments. Afterward, the EMD is able to extract no more than five IMFs and the residue for each 1 s EEG segment (Fig.2). For each one of these five IMFs, different parameters can be computed. The proposed parameters utilized in this work are the following:

- *Root Mean Square* (RMS),
- *Variance,*
- *Shannon entropy* [10],
- *Lempel-Ziv Complexity Measure* [11],
- *Central Frequency* (50 % of spectrum energy),
- *Maximum Frequency* (95 % of spectrum energy).

Some parameters were chosen since they are commonly used in BCI (RMS, variance), LZ quantifies the complexity of a signal analyzing its spatio-temporal patterns and was used for analyze EEG signals in other areas [12]. The central and maximum frequencies were used as descriptors of the band-with of each IMF. Entropy was used as a different metric, given that it measures the average amount of information from a measurement.

C. Feature Selection

A disadvantage arising at this point is that the feature vector that would enclose all the features calculated with the above parameters would be too large, i.e., each feature vector contains 180 parameters (5 IMFs x 6 parameters × 6 channels). Consequently, it is essential to do a feature selection in order to solve this curse-of-dimensionality inconvenience [13]. This selection is performed with a *stepwise method* based on the statistical parameter *Wilks' lambda* (WL). The WL measures the ratio of within-group variability respecting the total variability on the discriminator variables, and it is a measurement of the

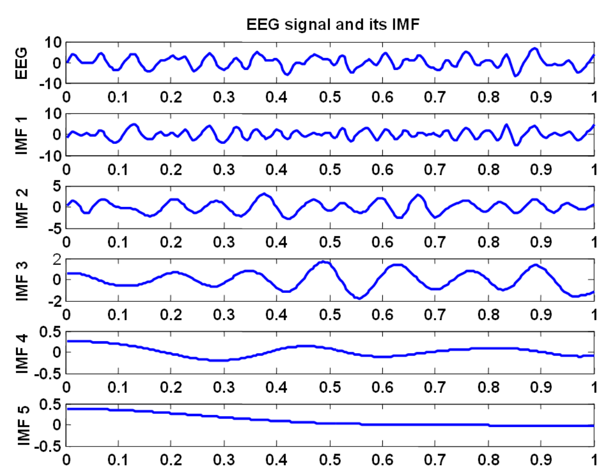

Fig. 2: Decomposition of an EEG segment of 1 s length in its IMF.

TABLE I
ACCURACY CLASSIFICATION OF MENTAL TASKS ON ALL SUBJECTS

Mental Tasks combination	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Mean	
	LD	**MLP**	**LD**	**MLP**	**LD**	**MLP**	**LD**	**MLP**	**LD**	**MLP**	**LD**	**MLP**	**LD**	**MLP**	**LD**	**MLP**
Base-Count	84,86%	86,67%	86,95%	80,84%	74,03%	70,14%	98,44%	95,94%	71,67%	71,67%	94,86%	81,53%	90,00%	79,17%	**85,83%**	**80,85%**
Base-Lett	95,42%	96,11%	85,28%	87,23%	78,89%	79,17%	97,81%	94,69%	84,54%	80,19%	95,70%	93,33%	99,45%	93,89%	**91,01%**	**89,23%**
Base-Math	97,64%	98,33%	99,17%	79,17%	83,61%	83,89%	96,56%	91,72%	84,16%	87,31%	96,25%	94,17%	100%	84,45%	**93,91%**	**88,43%**
Base-Rot	98,89%	96,95%	89,17%	85,28%	74,16%	72,50%	86,09%	78,59%	83,42%	81,76%	97,78%	94,17%	91,67%	94,45%	**88,74%**	**86,24%**
Lett-Count	98,33%	98,06%	92,50%	92,50%	80,14%	76,81%	96,25%	91,25%	85,37%	81,57%	92,50%	86,25%	100%	89,72%	**92,15%**	**88,02%**
Lett-Rot	99,44%	98,89%	97,78%	93,89%	87,78%	81,67%	94,53%	93,28%	95,47%	90,37%	96,11%	92,08%	100%	94,72%	**95,87%**	**92,13%**
Math-Count	98,89%	96,94%	88,61%	83,61%	82,50%	79,03%	90,94%	90,16%	82,13%	82,78%	90,42%	90,70%	99,17%	81,94%	**90,38%**	**86,45%**
Math-Lett	100%	97,64%	98,89%	82,23%	87,64%	86,11%	95%	90,31%	85,28%	81,02%	94,72%	90,00%	99,45%	96,39%	**94,45%**	**89,10%**
Math-Rot	100%	96,95%	96,12%	92,50%	87,36%	81,11%	91%	92,34%	92,96%	93,98%	94,30%	91,25%	100%	89,45%	**94,57%**	**91,08%**
Rot-Count	76,67%	70,83%	90,00%	88,06%	73,20%	70,70%	80,16%	82,03%	86,30%	82,59%	93,47%	87,08%	93,9%	92,50%	**84,81%**	**81,97%**
Average	**95,00%**	**93,74%**	**92,45%**	**86,53%**	**80,93%**	**78,11%**	**92,73%**	**90,03%**	**85,13%**	**83,32%**	**94,61%**	**90,06%**	**97,36%**	**89,67%**	**91,17%**	**87,35%**

LD: Linear Discriminate; MLP: Multi Layer Perceptron

importance of the variables. Therefore, the more important variables for the analysis should be selected, i.e. the variables that contribute with more information. Besides, the correlated variables are discarded in this process [14].

In a *p*-dimensional space constructed with *p* variables and with the matrixes $B_{p\ x\ p}$ and $W_{p\ x\ p}$ representing the square sum and cross products between groups and within-groups, respectively; the WL can be defined as the ratio between their determinants [14]:

$$WL = |W| / |B + W| \tag{5}$$

Then, the value of WL is transformed into the general multivariate statistical *F*, which allows contrasting significant differences between groups. A variable is accepted in the analysis, if *F* value is higher than 3.84 (*F* to enter) and, once included, the variable is rejected if its *F* value is smaller than 2.71 (*F* to exit). The stepwise method using the WL was implemented within each subject.

D. Classifier

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

1) Linear Classifier:

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

$$y = \mu_0 + \mu_1 X_1 + \mu_2 X_2 + ... + \mu_p X_p \tag{6}$$

where y is the output value of the discriminate function; μ_i are the coefficients of the discriminate function; X_i are the discriminate variables at each case and p is the number of variables in the analysis [14].

2) Nonlinear Classifier:

As a nonlinear classifier, neural networks were chosen. A multilayer perceptron (MLP) [15] with two hidden layers (with 10 and 5 neurons per layer, respectively) was implemented using Matlab®. In the output layer, one neuron per each mental state was utilized. The MLP was trained with Levenberg-Marquard backpropagation method, and an early stopping method was used to stop the training process, with 90% of the data used to train it, and the remaining 10% was used to validate it.

IV. RESULTS

The feature selection is an important issue in order to solve the curse of dimensionality [13] 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 task) in a *one-versus-one* classification scheme. Generally, only parameters from the foremost IMFs (1^{st} to 3^{rd}) were chosen in the analysis.

Table I shows the results obtained with the EMD in a *one-versus-one* scheme for each subject. These values are obtained using a 10-fold cross-validation repeated over four times, in order to obtain more accurate results. Therefore, the values shown in this table are the average values over results obtained in each cross-validation. In Fig.3, the average values of each subject and the mean over all subjects are presented.

V. DISCUSSION

The *one-versus-one* classification scheme shown in Table I indicate that the Subject 1 obtained, for all the classifiers, results above 80% for almost all the possible combinations. 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. Subject 3 showed the worst overall performance, with results lying between 70% and 87%. In the case of Subject 4, the results are similar to those of Subjects 1 and 2. Subject 5 attained values greater than 80% except in *Base-Count* combination (70%). Subjects 6 and 7 had an average performance of 94% and 97% with LD, respectively; and 89% with the MLP.

In the majority of mental tasks combinations, the best results were obtained using LD for all subjects. This fact is easy to see on 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% MLP. The ACC attained with this method were higher than the results documented in similar works with the same database [5-7]. In [5], power and asymmetry ratios of EEG bands were used with an average ACC of 86.5% on 4 subjects with an Elman neural network. In [6], autoregressive (AR) modelling and multilayer perceptron were used with ACC of up to 71%. The best result obtained in [7] was 72%, obtained with AR modelling and support vector machines.

It has been found that the WL parameter allows choosing the more suitable variables in the analysis, and to solve the curse-of-dimensionality (an important aspect of BCI applications [13]) by reducing the feature vector of 180 variables into a small number of 16±7 variables for each combination, which allows a better ACC. Generally, only the parameters from the foremost IMFs (1^{st} to 3^{rd}) were selected by *WL*, i.e., it is not necessary to extract all the IMFs of the signal. These IMFs contain, principally, frequencies ranging in α (8-13 Hz), β (14-30 Hz) and δ bands (> 30 Hz), those related with alertness and thinking states. The more chosen parameters were, in descendent order of importance, RMS, variance, LZ, entropy, maximum frequency and central frequency. The feature selection through the WL parameter accomplishes this objective, while being easy and fast to compute as well.

The EMD offer an advantage over other signal analysis methods, like spectral analysis or wavelet transform; since EMD is adaptive to the signal, whereas, in Fourier and wavelet transforms the basis are fixed. Hence, EMD allows extracting better features from non-stationary signals, such as EEG.

VI. Conclusions

In this work an alternative extraction features 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 different classifiers (LD and MLP) were employed.

This method allows attaining very high results in the ACC of mental tasks, obtaining performances greater than 90% for almost all subjects in a *one-versus-one* scheme using any classifier. The LD performs better than MLP with this method. The results were higher than those documented in similar works using the same database [5-7]. In these works ACC of 86%, 71% and 72% were presented in [5], [6] and [7], respectively. Although, these comparisons give us an idea of the performance of our method, it is not suited at all, due to differences in the subjects analyzed on database (all the subjects were used in this work), different classifiers, and different training and validation sets used in those works.

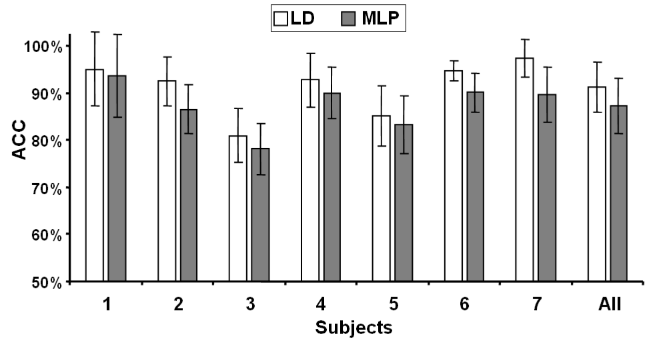

Fig. 3: Average values of each subject and the mean over all subjects

Acknowledgment

The authors want to thanks to Dr. C.W. Anderson of Colorado State University for make available the EEG database used and to P. Flandrin *et al.* for the EMD toolbox.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller , and T. M. Vaughan, "Brain-computer interfaces for communication and control" Electroen. Clin. Neuro, vol. 113, no. 6, pp. 767-791, June 2002.

[2] D. J. McFarland, C. W. Anderson, K-R. Müller, A. Schlögl, and D. J. Krusienski, "BCI Meeting 2005—Workshop on BCI Signal Processing: Feature Extraction and Translation" IEEE Trans. Neural Syst. Rehabil. Eng., vol. 14, no. 2, pp. 135-138, June 2006.

[3] N. E. Huang, Z. Shen, S. R. Long, M. L.Wu, H. H. Shih, Q. Zheng, N. C.Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis," Proc. R. Soc. London A, vol. 454, pp. 903–995, 1998.

[4] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings," IEEE Trans. Biomed. Eng., vol. 37, no. 12, pp. 1209–1214, Dec. 1990.

[5] R. Palaniappan, "Utilizing Gamma Band to Improve Mental Task Based Brain-Computer Interface Design" IEEE Trans. Neural Syst. Rehabil. Eng., vol. 14, no. 3, pp. 299-303, Sep. 2006.

[6] C. W. Anderson and Z. Sijercic "Classification of EEG signals from four subjects during five mental tasks" Solving Eng. Problems with Neural Networks: Proc. Int. Conf. on Eng. Appl. of Neural Net. (EANN'96), London, 17th-19th June, 1996.

[7] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification" IEEE Trans. Neural Syst. Rehabil. Eng. 11 141–4. 2003.

[8] http://www.cs.colostate.edu/eeg/index.html#Data

[9] EMD Matlab toolbox: http://perso.ens-lyon.fr/patrick.flandrin/emd.html.

[10] C. E. Shannon "A Mathematical Theory of Communication" AT&T Tech. J., vol. 27, pp. 379–423, 623–656, July, October, 1948.

[11] A. Lempel and J. Ziv,"On the complexity of finite sequences", IEEE Trans. Inform. Theory, vol. IT-22, pp. 75-81, 1976.

[12] X. S. Zhang, R. J. Roy, and E. W. Jensen, "EEG complexity as a measure of depth anesthesia for patients," IEEE Trans. Biomed. Eng., vol. 48, pp. 1424-1433, Dec 2001.

[13] F. Lotte, M. Congedo, A. L´ecuyer, F. Lamarche, and B. Arnaldi "A review of classification algorithms for EEG-based brain–computer interfaces" J. Neural Eng. 4 (2007) R1–R13, January 2007.

[14] J. Gil Flores, E. García Giménez, and G. Rodríguez Gomez, "Books of Statistics Nº12: Discriminate Analysis" (in Spanish). Ed. La Muralla S.A. and Ed. Hespérides S.I. 2001. pp. 31-57.

[15] S. Haykin "Neural Networks A Comprehensive Foundation"– 2º Edition. New Jersey, USA. - Ed. Pretince Hall, 1999. ch. 4