2011 IEEE International Conference on Rehabilitation Robotics
Rehab Week Zurich, ETH Zurich Science City, Switzerland, June 29 - July 1, 2011

Nonlinear and nonstationary framework for feature extraction and classification of motor imagery

Dalila Trad
UVSQ, LISV,
10/12 Avenue de l'Europe,
78140 Vélizy, France
and
University of Tunis, UTIC,
5, avenue Taha Hussein,
B.P. 56 Bab Menara 1008 - Tunis Tunisia
Email: d.trad@esiee.fr,

Tarik Al-ani
UVSQ, LISV,
10/12 Avenue de l'Europe,
78140 Vélizy, France
and
ESIEE-Paris, Dep. Informatique,
Cité Descartes-BP 99,
93162 Noisy-Le-Grand, France
Email: t.alani@esiee.fr
Telephone: +33 1 45926598
Fax: +33 1 45926699

Eric Monacelli
Stéphane Delaplace
UVSQ, LISV,
10/12 Avenue de l'Europe,
78140 Vélizy, France
Email: eric.monacelli@uvsq.fr

Mohamed Jemni
University of Tunis, UTIC,
5, avenue Taha Hussein,
B.P. 56 Bab Menara 1008 - Tunis Tunisia
Email: mohamed.jemni@fst.rnu.tn

***Abstract*—In this work we investigate a nonlinear approach for feature extraction of *Electroencephalogram* (*EEG*) signals in order to classify motor imagery for *Brain Computer Interface* (*BCI*). This approach is based on the *Empirical Mode Decomposition* (*EMD*) and *band power* (*BP*). The *EMD* method is a data-driven technique to analyze non-stationary and nonlinear signals. It generates a set of stationary time series called *Intrinsic Mode Functions* (*IMF*) to represent the original data. These *IMF*s are analyzed with the *power spectral density* (*PSD*) to study the active frequency range correspond to the motor imagery for each subject. Then, the band power is computed within a certain frequency range in the channels. Finally, the data is reconstructed with only the specific *IMF*s and then the band power is employed on the new database. The classification of motor imagery was performed by using two classifiers, *Linear Discriminant Analysis* (*LDA*) and *Hidden Markov Models* (*HMM*s). The results obtained show that the *EMD* method allows the most reliable features to be extracted from *EEG* and that the classification rate obtained is higher and better than using only the direct *BP* approach.**

I. Introduction

Brain Computer Interfaces (*BCI*) is a direct communication pathway between a brain and an external device. The major goal of the *BCI* research is to develop systems which help disabled users to communicate with other people in order to control artificial limbs or their environment [1], [2]. A *BCI* system is represented as a system in a continuous closed loop, generally composed of six steps, Figure 1: 1. Brain activity measurement, 2. Preprocessing, 3. Feature Extraction, 4. Classification, 5. Translation into a command and Feedback [3].

One major challenge of *BCI* is thus to extract reliable information (features) from noisy *EEG* data, i.e. step 3. These features can then be used in step 4 in order to classify the user's mental state. Physiological studies [4], [5], showed that the rhythms mu (8-12 Hz) and beta (13-30 Hz) are the main relevant information for discriminating motor activity. A common approach in *BCI* is thus to extract the *band powers* of the rhythms mu and beta from the *EEG* signal and use them as a classification features. Several common *band power* techniques were employed in the *BCI* literature. Herman et al. [6] demonstrated that the Yule and Welch *PSD* approaches, mainly dominate the other studied approaches. These approaches are based essentially on some linearity and stationarity hypothesis such as the use of *fast Fourier transform* (*FFT*) spectrum in short time of a segment of data. The accuracy of the *FFT* calculation is closely related to the choice of the duration of the signal segment [7]. However, the nature of the *EEG* signal is nonstationary and nonlinear [8]. The main nonstationary and nonlinear feature extraction technique is the *Wavelet Transform* (*WT*) [9]. Although this approach is more effective than the *FFT*, it shows at the same time much bigger ambiguity in signal decomposition. Huang et al. [10] proposed a more fairly recent technique called the *empirical mode decomposition* (*EMD*) for nonlinear and nonstationary

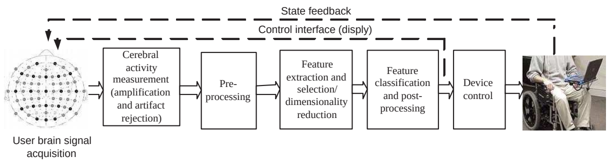

Fig. 1. General functional model of a *BCI* System.

time series data. The *EMD* is a data driven approach (i.e. one does not need to define a mother wavelet beforehand) that can be used to decompose adaptively a signal into a finite number of mono-component signals, which are known as *intrinsic mode functions* (*IMF*s) or *modes*. It considers signals at their local oscillations, but they are not necessarily considered in the sense of Fourier harmonics. Their extraction is non-linear, but their recombination for exact reconstruction of the signal is linear. The *IMF*s admit well-behaved *Hilbert transforms* (*HT*) [11] and they satisfy the following properties: they are symmetric, different *IMFs* yield different instantaneous local frequencies as functions of time that give sharp identifications of embedded structures. The decomposition is done linearly or non-linearly depending on the data. This complete and almost orthogonal decomposition is empirically realized by identifying the physical local characteristic time scales intrinsic to these data, which is the lapse between successive extrema.

The *EMD* was recently applied to *EEG* analysis such as detection of synchronisation [12] and motor imagery, [13], [14] where the Fourier spectra corresponding to the the rhythms mu and beta were constructed by the *EMD*.

Transient neural assemblies mediated by synchrony in particular frequency ranges are thought to underlie cognition. We propose a new approach to the detection of synchronisation in EEG

In this work, we apply first the *EMD* on the *EEG* signals and then we apply the Welch-based *band power* for feature extraction in order to extract the reliable information of *EEG* corresponding to some motor imagery tasks. Based on these features, the classification of the mental tasks was done using two classifiers: linear discriminate classifier (*LDC*) based on linear discriminant analysis (*LDA*) [15] and a nonlinear classifier known as *hidden Markov models* (*HMMs*) [16].

II. Data bases & Methods

A. EEG Data

Two motor imagery *EEG* data corresponding to two subjects were used in this work.

- For subject 1, the *EEG* data corresponds to an experiment with four sessions "run1234" acquired by Guger Technologies [17]. Each session contains 40 trials: 20 trails for of right hand movement imagination and 20 trials for left hand movement imagination, where each trial contains 2048 samples. The timing of this experiment is shown in Figure 2. The mechanism is described as follows: after two seconds a warning stimulus was given of a 'beep'. From second 3 until second 4.25 an arrow pointing left or right hand was shown on the monitor after that the subjects were instructed to imagine a left or right hand movement depending on the direction of the arrow.
- For subject 2, the *EEG* data is recorded in our Department using the biosignal amplifier "g.USBamp-Gtec" [18]. One subject (female aged 24 years) was instructed to imagine a left or right hand movement depending on the direction of the arrow. The same preceding experiment and data structure were obtained.

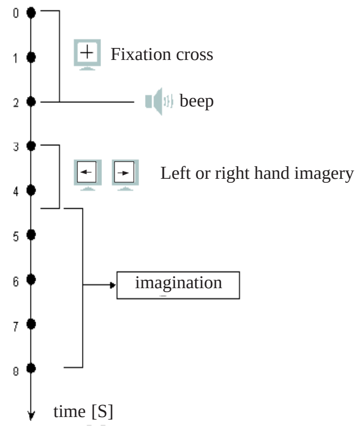

Fig. 2. Paradigm: Timing of one trial in the experiment.

The experiment data were sampled at 256 Hz and filtered in the range of 0.5 and 30Hz. A notch filter was used to suppress the 50Hz power line interference. Two bipolar recordings overlying the left and right sensorimotor area were obtained by two electrodes C3 and C4 placed according to the international 10/20 system [19].

B. Feature extraction

In brain-computer interface, feature extraction has an important role. In order to extract the relevant features of the *EEG* signal, we employed a feature extraction method based on *EMD* and *BP* in order to recognize the left or right motor imageries. The new feature extraction scheme is presented in Figure 3.

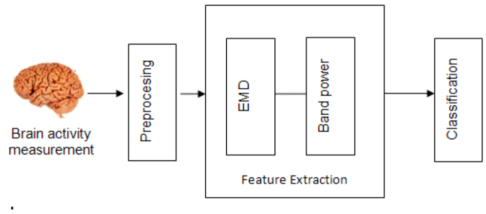

Fig. 3. Our feature extraction scheme.

1) Empirical mode decomposition (EMD) approach: The traditional *EMD* was recently proposed [10] as an adaptive time-frequency data analysis method. It is defined by an algorithm based on an empirical framework. In most cases, the studies (performance, analysis,...) carried out on the *EMD* are done with extensive digital simulations in controlled conditions [10]. Despite the lack of theoretical formalism, this algorithm showed its capacity to analyze the signals. Using a new formulation for *EMD* based on constrained optimization, the results of [20] were very similar to those obtained with the traditional *EMD* algorithm.

The basic *EMD* is defined by a process called *sifting* to break down any multimodal signal to a sum of basis components called *intrinsic mode functions* (*IMFs*). The *IMFs* are zero-mean AM-FM signals which must satisfy two conditions: the first one is that the number of extrema and that of zero-crossing must differ at most by one; the second one is that the mean value between the upper and lower envelopes are equal to zero at any point. Conceptually, the establishment of this method is quite simple: one needs to consider a signal at its local oscillation level, remove the fastest oscillation and iterate the process on the residue considered as a new signal. At the end of the sifting processes, a given signal $x(t)$ can be written as a sum of a finite number of *IMFs*, $I_m(t), m = 1,\ 2, ..., M$, and a final residue $r_M(t)$:

$$x(t) = \sum_{m=1}^{M} I_m(t) + r_M(t).$$

The decomposition is stopped at step M, if either the residue $r_M(t)$ is a mono-component signal or has only 2 extrema [10]. The *stopping criterion* must be set to ensure that the obtained signal satisfies the properties of an *IMF* while limiting the number of iterations. For more details about the different steps of the *sifting* process for the calculation of the IMF_i as well as the *stopping criterion* definition see [10]. Since the decomposition into *IMFs* is based on the local characteristic time scale of the data, it applies to nonlinear and non-stationary processes.

*2) Band powers (*BP*):* The features may be extracted from the *EEG* signals by estimating the power distribution of the *EEG* in predefined frequency bands. In general, the band power is estimated by digitally bandpass filtering the data, squaring and averaging over consecutive samples according to a given window size. Pfurtscheller et al. [21] used the *BP* and demonstrated that for each subject, different frequency components in the mu and beta bands were found which provided best discrimination between left and right hand movement imagination. These frequency bands varied between 9 and 14 Hz and between 18 and 26 Hz.

In this work, we propose a direct nonlinear approach to extract the more relevant *IMF*s corresponding to the different frequency components in the mu and beta bands and then obtain the Welch-based *BP* and use them as features for mental task classification.

We applied the *EMD* method on the *EEG* data defined in section II-A. The *EEG* data for each subject are composed of 80 trials corresponding to left hand movement imaginations (C4) and 80 trials corresponding to right hand movement imaginations (C3). Figure 4 shows the result of one-trial *EMD* decomposition for subject 2. Each channel is decomposed into 10 *IMF*s and one residue.

To analyze the different characteristics of each *IMF*, we applied the Welch's method [22]. This method estimates the *PSD*, it was applied to each *IMF* to calculate and find the active frequency bands such as the mu and beta rhythms. Figure 5 shows the *PSD* in each *IMF* in the two channels C3 and C4.

We can notice that the characteristics of the active frequency bands corresponding to mu and beta are located only in *IMF*1, *IMF*2 on C3 and C4. Concerning subject 1, the active

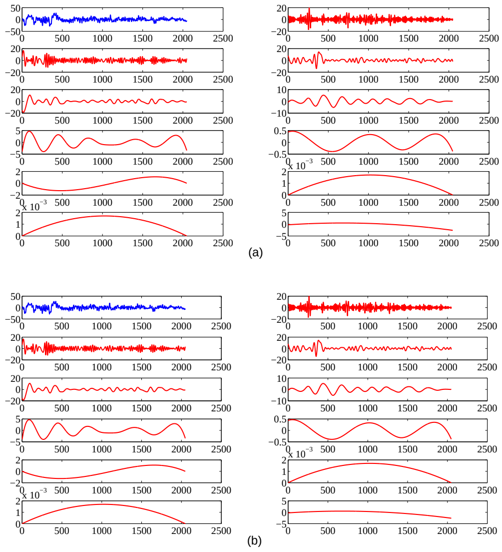

Fig. 4. (a) From top-left to down-right: the raw signal, the ten *IMF*s and the residue in channel C4 of subject 2 as a function of sample index. (b) from top-left to down-right: the raw signal, the ten *IMF*s and the residue in channel C3 of subject 2 as a function of sample index.

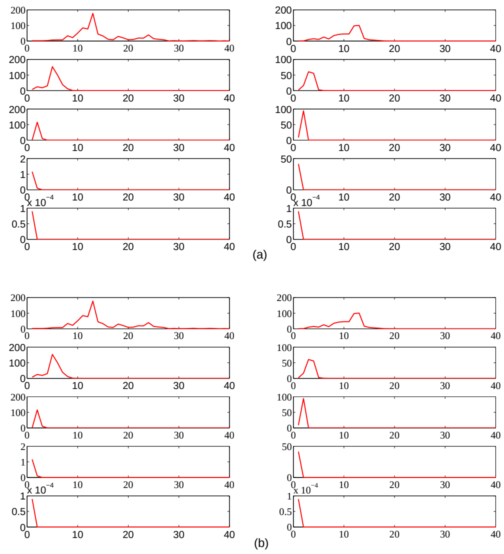

Fig. 5. (a) PSD (dB/Hz) vs. frequency (Hz) of each *IMF* in (a) channel C4 and (b) channel C3 of subject 2.

frequency bands are located only in *IMF*1, *IMF*2 and *IMF*3 on C3 and C4. Therefore, the new signal is reconstructed by keeping only the two first *IMF*s for subject 2 and only three first *IMF*s for subject 1. Then, band power was applied for the new signal.

C. Classification

Two different classifiers were implemented to classify the different motor imagery (imagination of right or left hand movement):

1) *LDC*: The *LDC* is a linear discriminant classifier based on *LDA*. It has been used with success in many of *BCI*s such as motor imagery [4], [5]. The idea of *LDA* [15] is to find a weight vector $\mathbf{W}$ so that two projected clusters c1 and c2 of N1 and N2 training feature vectors on $\mathbf{W}$ can be well separated from each other by hyperplanes while keeping small variance of each cluster. The parameters are obtained with a learning algorithm from a set of training data. Fisher introduced a method that reduces the dimensionality before classification [15], [23]. The dimension reduction is done by projecting the input data $\mathbf{X}$ onto a value $\mathbf{y}$ with adjustable weights vector $\mathbf{W}$:

$$y = \mathbf{W}^T \mathbf{X}.$$

The separating hyperplane is obtained by seeking the projection that maximizes the distance between the two classes' means and minimizes the between variance.

2) *HMMs*: This method is very efficient nonlinear technique used for the classification of time series [16]. it necessitates two stages: a training stage where the stochastic process models are estimated through extensive observation corpus and decoding or detection stage where the model may be used off/on-line to obtain the likelihoods of the given test sequence evaluated by each model [24], [25]. A *HMM* is defined by the following compact notation to indicate the complete parameter set of the model $\lambda = (\mathbf{\Pi}, \mathbf{A}, \mathbf{B})$, where $\mathbf{\Pi}$, $\mathbf{A}$ and $\mathbf{B}$ are the initial state distribution vector, matrix of state transition probabilities and the set of the observation probability distribution in each state, respectively [16]. This set of parameters is defined by

$$\mathbf{\Pi} = [\pi_1, \pi_2, \; ..., \pi_N], \;\; \pi_i = P(q_1 = s_i),$$

$$\mathbf{A} = [a_{ij}], \;\; a_{ij} = P(q_{t+1} = s_j | q_t = s_i).$$

Where $1 \leq i, j \leq N$, s_i, $s_j \in S$, $S = \{s_1, s_2, \; ..., s_N\}, t \in \{1, 2, \; ..., T\}$. The observation at time (or index) t, $\mathbf{O}_t$, is considered in this paper as continuous $\mathbf{O}_t \in \mathbb{R}^K$. For a continuous observation, the state conditional probability of the observation $b_i(\mathbf{O}_t)$ may be defined by a finite mixture of any log-concave or elliptically symmetric probability density function (pdf), e.g. Gaussian pdf, with state conditional observation mean vector $\boldsymbol{\mu}_i$ and state conditional observation covariance matrix $\mathbf{\Sigma}_i$. In this paper we consider only a single Gaussian pdf, so $\mathbf{B}$ may be defined as $\mathbf{B} = \{\boldsymbol{\mu}_i, \; \mathbf{\Sigma}_i\}, \; i = 1, 2, \; ..., N$. At each instant of time t, the model is in one of the states i, $1 \leq i \leq N$. It outputs $\mathbf{O}_t$ according to a density function $b_i(\mathbf{O}_t)$ and then jumps to state j, $1 \leq j \leq N$ with probability a_{ij}. The state transition matrix defines the structure of the *HMM* [16]. The model λ may be obtained off-line by a given training procedure. In practice, given the observation sequence $O = \{\mathbf{O}_1, \mathbf{O}_2, \; ..., \mathbf{O}_T\}$, and a model λ, the *HMMs* need three fundamental problems to be solved:

a) How to calculate the likelihood $P(O|\lambda)$? The solution to this problem provides a score of how O belongs to λ.
b) How to determine the most likely state sequence that corresponds to O? The solution to this problem provides the sequence of the hidden states corresponding to the given observation sequence O.
c) How to adjust the model λ in order to maximize $P(O|\lambda)$? This is the problem of estimating the model parameters given a corpus of training observations sequences.

Problems 1 and 2 are solved in the decoding or detection stage using the forward or the Viterbi algorithms [16], while problem 3 is solved during the training phase using either a conventional algorithm such as the Baum-Welch algorithm [16].
Our training scheme is based on Baum-Welch algorithm and the *Bayesian Inference Criterion* (*BIC*) [26] [27]. This scheme makes the training procedure independent of the initialization problem and the a priori knowledge of the number of states in each *HMM* needed in the Baum-Welch training algorithm.

D. Results

Table I shows the classification results based on the two feature extraction methods (*BP* and *EMD* + *BP*) and by using the two classification approaches *HMMs* and *LDA*. For

TABLE I
RECOGNITION PERCENTAGE RATES USING *HMMs* AND *LDA* CLASSIFICATION (%) OF MOTOR IMAGERY (LEFT AND RIGHT HAND MOUVEMENT).

Subject	subject 1		subject 2	
Feature extraction method	*BP*	*EMD + BP*	*BP*	*EMD+BP*
HMMs **classifier**	87.7%	91.25%	62.5%	76.25%
LDA **classifier**	69.76%	73.1%	78.29%	83.12%

each subject, the *EEG* data contains 160 trials and each trial lasts 8 seconds as shown in Figure 2 (a set of 80 trials for left hand movement imaginations and a set of 80 trials for right hand movement imaginations). Each set of movement

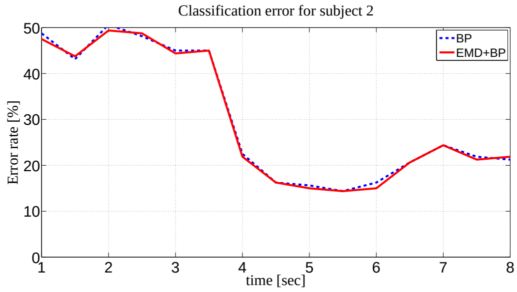

Fig. 6. Classification rate as a function of time (sec.) for right and left hand movement imagery using *LDA* and feature extraction : *BP* and *EMD+BP* of subject 2.

imagination data was divided into two subsets for each mental task movement (40 trials for *LDA*/*HMMs* training and 40 trials for test). In these subsets, we considered only the imagination period: 4 seconds to 8 seconds (see Figure 2).

For the *LDA*, the recognition percentage rates shown in Table I represent the average recognition rates between 4s and 8s calculated on the 40 test trials. Figure 6 shows an example of one test-trial classification rate as a function of time (sec.) for right and left hand movement imagery using the feature extraction methods *BP* and *EMD*+*BP* of subject 2. In this example, the classification results show that at the beginning of the trial the error is around 50%. After second 4 (arrow is shown to the subject on the screen at second 3, see Figure 2) the error drops down. In Figure 6, the averages of classification error for subject 2 in the imagination phase are around 21,71% and 16,88% by using *BP* and *EMD*+*BP* respectly. This means that the data set can be classified with an accuracy of about 78,29% and 83,12% respectively (see Table I).

For the *HMMs*, the recognition percentage rates hown in Table I correspond to the average of recognition percentage rates in the diagonal of confusion matrices shown in Table II. It can be seen that the classification results with the new feature extraction method combining *EMD* and *BP* is better than using only *BP* method. It can be seen also that the *HMMs* give better results than the *LDA* in all the cases.

TABLE II

CONFUSION MATRICES FOR LEFT AND RIGHT HAND MOVEMENTS USING *HMMs*. FROM TOP TO DOWN: USING *BP* METHOD FOR SUBJECT 1, USING *BP+EMD* METHOD FOR SUBJECT 1, USING *BP* METHOD FOR SUBJECT 2 AND USING *BP+EMD* METHOD FOR SUBJECT 2.

	Right	Left
Right	85.0%	15.0%
Left	10.0%	90.0%

	Right	Left
Right	85.0%	15.0%
Left	2.50%	97.5%

	Right	Left
Right	72.5%	27.5%
Left	47.5%	52.5%

	Right	Left
Right	72.5%	27.5%
Left	20.0%	80.0%

III. CONCLUSION

In this work, a feature extraction method based on the Empirical Mode Decomposition (*EMD*) and the *band power* (*BP*) is proposed to keep only the active frequency band powers corresponding to mu and beta rhythms in *BCI*-related mental task *EEG* signals. The feature extraction process is done in three-stages: in the first stage applies the *EMD* on the raw *EEG* signals to obtain the *Intrinsic Mode Functions* (*IMF*). The second stage reconstructs the relevant signal by keeping only the *IMF*s which contain the active frequency bands. The third stage calculates the *BP* of the active frequencies in the relevant signal. To evaluate this approach, two classifiers were employed: a linear classifier based on *linear discriminant analysis* (*LDA*) and nonlinear classifier based on *hidden Markov models* (*HMMs*). The proposed feature extraction approach gives better classification results for the two classifiers and seems promising. However, to use the *HMMs* classifier, some care should be taken into account such as the necessity use of a large data base for training or the use of a good discriminative training algorithm to train the *HMMs* [28].

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, J. McFarland D., G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control", *Clinical Neurophys*, vol. 113, 2002, pp. 767-791.

[2] J. R. Wolpaw, "Brain-computer interfaces (BCI's) for communication and control: Current status," *2nd Int. BCI Workshop Training Course*, 2004, pp. 29-32.

[3] S. G. Mason and G.E. Birch, "A general framework for brain-computer interface design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*,11(1):70-85, 2003.

[4] G. Pfurtscheller, EEG Event-related Desynchronisation (ERD) and Event-related Synchronisation (ERS). In Niedermeyer, pp. 958-967, E., Lopes da Silva, F.H. (Eds.), Electroencephalography. Basic Principles, Clinical Applications, and Related Fields, forth ed., Williams and Wilkins, Baltimore.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*," vol. 89, no. 7, pp. 1123-1134, 2001.

[6] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 16, no. 4, 2008, pp. 317-326.

[7] T. Penzel and R. Conradt, "Computer based sleep recording and analysis", Sleep Medicine Reviews, vol. 4, no. 2, pp. 131-148, 2000.

[8] M. Lo, P. Tsai, P. Lin, C. Lin, and Y. L. Hsin, "The Nonlinear and non-stationary Properties in EEG Signals: Probing the Complex Fluctuations by Hilbert-Huang Transform," presented at Advances in Adaptive Data Analysis, pp.461-482, 2009.

[9] S. G. Mallat, "A theory for multiresolution signal decomposition : the wavelet representation," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693, July, 1989.

[10] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. of the Royal Society, vol. 454, no. 1971, pp. 903-995, 1998.

[11] D. Gabor, "Theory of communication," Proc. IEE, vol. 93 (III), pp. 429-457, 1946.

[12] C. M. Sweeney-Reed and S. J. Nasuto, "A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition", J Comput Neurosci vol. 23, pp. 79111, 2007.

[13] G. Xiaojing, W. Xiaopei, and Z. Dexiang, "Motor Imagery EEG Detection by Empirical Mode Decomposition," *Conf Proc IEEE IJCNN 2008*, pp. 2619-2622, 2008.

[14] L. Wang, X. Guizhi, J. Wang, S. Yang, and W. Yan, "Application of Hilbert-Huang transform for the study of motor imagery tasks," *Conf Proc IEEE Eng Med Biol Soc.*, pp. 3848-3851 , 2008.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Recognition," 2nd. ed. (New York: Wiley-Interscience), 2001.

[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected application in speech recognition", *Proc. IEEE*, vol. 77, no. 2, February 1989, pp. 267-296.

[17] Guger Technologies, "BSanalyze advanced biosignal processing and analysis version 3.07a,". *http://www.gtec.at/*.

[18] *http://www.gtec.at/*

[19] E. Niedermeyer, F. Lopes da Silva, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields," Lippincott Williams & Wilkins, 2004 ISBN 0781751268, 9780781751261.

[20] S. Meignen and V. Perrier, "A New Formulation for Empirical Mode Decomposition Based on Constrained Optimization," Signal Processing Letters, IEEE, vol. 14, pp. 932-35, 2007.

[21] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," Electroenceph clin Neurophysiol, vol. 103, pp. 642-651, 1997.
[22] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," IEEE Transactions on Audio Electroacoustics, vol. AU-15, pp. 7073, 1967.
[23] C. M. Bishop, "Neural Networks for pattern Recognition," Clarendon Press,Oxford,1995.
[24] T. Al-ani and D. Trad, "Signal Processing and Classification Approaches for Brain-computer Interface", In : Intelligent and Biosensors, Edited by Vernon S. Somerset ISBN 978-953-7619-58-9, pp. 25-66, Publishing date: January 2010.
[25] D. Trad, S. Helmy, T. Al-ani, and S. Delaplace, "Hidden Markov Models and Support Vector Machines applied to P300 Based Brain-Computer Interface - a Comparative study", ICTA 09, Hammamet, 7-9 Mai, 2009.
[26] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, Vol. 6, no. 2, pp. 461-464, 1978.
[27] S. Helmy, T. Al-ani, Y. Hamam, and E. El-madbouly, "P300 Based Brain-Computer Interface Using Hidden Markov Models," ISSNIP 2008, Sydney, 15-18 Dec., 2008.
[28] T. Al-ani, "Hidden Markov models in dynamic system modelling and diagnosis, Hidden Markov Models, Theory and Applications," ISBN: 978-953-307-208-1, InTech, In Press.