A Novel Feature of the EEG based Motor Imagery BCI System: Degree of Imagery

Ching-An Cheng[1], *Han-Pang Huang*[2]*, Member, IEEE, *Yi-Hung Liu*[3], Member, IEEE
[1], [2*]*Department of Mechanical Engineering*
National Taiwan University
Taipei, Taiwan
E-mail: hanpang@ntu.edu.tw
[4] *Department of Mechanical Engineering*
Chung Yuan Christian University
Taoyuan County, Taiwan

***Abstract*—Motor imagery recognition has been considered an important topic in the brain-computer interface (BCI) community. Due to noises and artifacts in signals, how to gain satisfactory classification accuracy is still a critical issue. We propose in this paper a novel feature to address this issue. The method consists of three steps. Firstly, EEG signals from different electrodes are transformed by Time-Frequency Analysis method, in this paper Hilbert-Huang Transform. A set of features, Degree of Imagery (DOI) are then extracted from the spectrums by the proposed feature extraction method. The features can effectively represent the event-related-desynchronization (ERD) during motor imagery. Experiments on the BCI 2003 competition dataset III indicate that our method achieves better classification accuracy and higher mutual information (MI) than other researches using the same dataset and with low computational time, which is capable of real-time usage.**

***Keywords*—motor imagery, degree of imagery (DOI), brain-computer interface Hilbert-Huang transform, mutual information, event related desynchronization,**

I. INTRODUCTION

The Brain Computer Interface (BCI) is an interface technique between human and computer which can help severely motor-disabled patients to communicate and control the environment. Usually the EEG signals are preprocessed to improve performance. High-Pass Filter and Low-Pass Filter are used to eliminate the noise and artifacts (EOG or EMG). However, the artifacts and the desired EEG are usually in the same frequency range. Various feature extraction methods have been proposed, such as ICA and PCA, etc. However the results are not satisfactory due to the non-stationary characteristics of EEG and the limited knowledge to the brain function. Therefore, how to extract distinguishing features from EEG becomes critical for motor imagery recognition.

During the motor imagery, the energy in the mu band (8~12Hz) at the contralateral hemisphere sensorimotor representative areas declines, Event-Related Desynchronization (ERD), and increases at the ipsilateral hemisphere, Event-Related Synchronization (ERS). Many methods has been proposed in the decade, and one branch of them use the Time-Frequency Analysis (TFA) extract the feature of motor imagery. Mostly, the raw EEG data is transformed into frequency domain by methods like Fourier Transform and Wavelet Transform. Recently, some researches [1-3] used the Hilbert-Huang Transform (HHT) [4], however the results seems not good enough comparing to others. The Hilbert spectrum in HHT provides an adaptive analysis method and a better transient resolution than that of Fourier Transform and Wavelet Transform. However the question lies on whether a better transient resolution provides us a better view for imaginary movement detection. Hilbert spectrum is more discrete than the frequency spectrum of Fourier Transform and Wavelet Transform. On the other hand, the ERD effect is a macroscopic phenomenon to time, while Hilbert spectrum shows as a microscopic view. Thus some modification must be made to suit for the classification of imaginary movement when using HHT. In this paper, we devise a feature, Degree of Imagery (DOI), based on HHT. This method can effectively detect the ERD during motor imagery, thereby improving the classification performance.

In the previous researches, mostly only either the accuracy or the mutual information (MI) [5] of the classification algorithm is concerned. To implement an online BCI system, the calculation speed is also an important factor. With exhaustive computational time, the system would no longer be practical, even when the accuracy of the system is quite high.

In this paper, the BCI competition 2003 data set III is adopted to compare the performance of the proposed method and the results of other researches, because the data set is well recognized with validity in the research in BCI. It's also important to mention that some of the researches focus on the results of the Offline Classification, i.e. only one outcome per trial; and some other researches implement their algorithms to be Online Simulation, i.e. using the data set to simulate the online classifying circumstance. Therefore, the experimental results of this paper would perform in the both condition in order to compare the results with different paper using the same data set. The results show a high accuracy and MI system, using the novel feature, Degree of Imagery (DOI), with a rather low computational time.

This paper is organized as follows. In Section 2, we first describe the EEG data set used in this work. Section 3 introduces our method and the feature, DOI, in detail. Experimental results and the comparison of the previous researches are presented in Section 4. Discussion and Conclusion are drawn in Section 5.

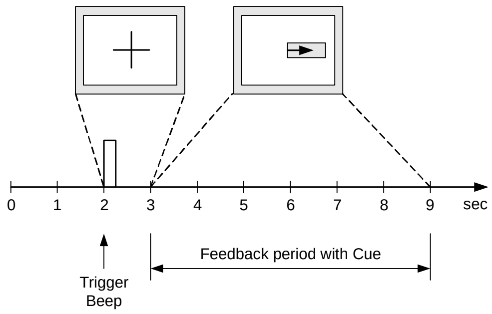


Figure 1. The paradigm of the Graz BCI competition II, 2003.

II. DATA DESCRIPTION

The data used in this work are those from the Graz BCI competition II dataset III, 2003 [6]. The data were from a normal subject, a 25 years old female, during a feedback session. The session consists of 7 runs with 40 trials each. Each trial is of 9 s, and is illustrated in the paradigm as Figure 1. In the period of 0-2 s, the subject was asked to relax. At $t = 2$ s, an acoustic stimulus indicates that a motor imaginary task is ready to start, and then the symbol "+" is displayed on the screen for 1 s. Next, an arrow (left or right) is displayed as a cue, which lasts for 6s ($t = 3$ $t = 9$ to). During the feedback period, the subject was asked to imagine a right or left hand movement. During the 9s-length trial, the EEG signals, recorded from C3, Cz, and C4, where Cz is a reference, were collected with a sampling rate of 128 Hz and filtered between 0.5 and 30 Hz.

III. PROPOSED METHODS

A. *Scheme*

The classification scheme is described in Figure 2. In the training stage, first, the training dataset of EEG signal is transformed by Time-Frequency Analysis (TFA) method to get the time-frequency patterns. Second, the proposed feature selection method, Degree of Imagery (DOI), is applied on the spectrum, and the parameters are determined by the 10-fold cross validation. Finally, the parameters of the classifier, SVM, are trained by the 10-fold cross validation to obtain the parameters with the highest accuracy. In the testing stage, the testing dataset is processed the same way as in the training stage, however, with the fixed parameters obtained previously. The mode the output can be chosen either to be Online Simulation or Offline Classification, in order to compare and correspond to the performance of other research using the same dataset.

Ideally, the TFA method in the scheme is arbitrary. In this paper, the TFA method is chosen to be HHT (for the implantation details, please refer to [4]), and Gabor Transform to test the proposed feature selection method. The reason to choose HHT is that it's less frequently adopted in the BCI motor imagery tasks in comparison to other TFA techniques and the adaptive nature. Gabor Transform serves as the representative of the traditional TFA methods.

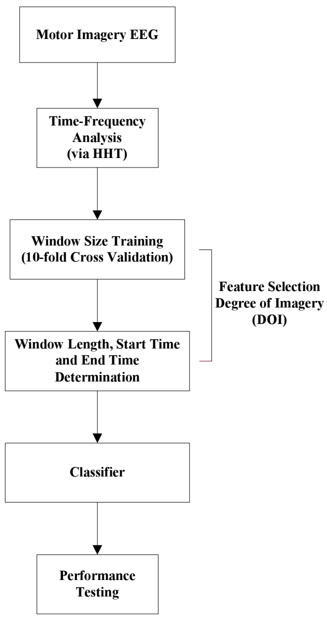


Figure 2. The scheme of the proposed method

B. *Improving the Computational Speed of HHT*

While HHT appears to be effective in detecting ERD, it is computationally expensive. A large number of iterations will be required in generating each IMF, making the EMD process time-consuming. As defined, the EMD process is stopped until a monotonic function (the residue) is obtained. However, one does not need to accomplish the whole EMD process in practice. In other words, even if all IMFs of an EEG signal are found, not all of them will be useful for the detection of ERD.

First, the sum of the energy of the first three IMFs contributes more than 80% of the total energy of the raw EEG signal, and the other part of the energy is almost equally distributed over the rest, shown in Figure 3.

Second, Figure 4 gives another view of how different IMFs take part in an EEG signal. Figure 4 is the averaged Hilbert spectrum of 140 different EEG signals of imaginary movement, where the numbers of right-hand and left-hand imaginary movements are equal. It shows that the first IMF stochastically well distributed over the frequency range from DC to 32Hz, the second IMF focuses on the frequency range from 8 to 13Hz, the third IMF occupies the frequency range from DC to 10Hz, and the frequency ranges of the other IMFs mainly lie near DC.

Third, the previous stochastic conclusion is on the EEG signals sampled in 128 Hz, however the outcomes of HHT differs when different sampling frequencies and interpolation method is applied. Stochastically, HHT acts as dyadic filter banks [7]. Since the Nyquist frequency is 64 Hz, it can be found that from the first to third the filter banks stochastically almost covers the entire effective frequency range. Next, we introduce how to extract features from the Hilbert spectrum.

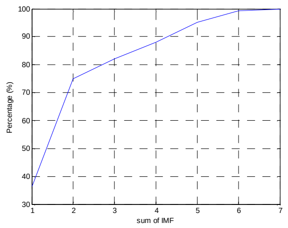


Figure 3. This figure summarizes the energy contribution of the IMFs. The contribution of different IMFs of EEG signals. The label 1 in the x axis denotes the first IMF, and label 2 denotes the percentage of the energy contributed by the 1st and the 2nd IMFs and so on.

C. *Feature Extraction: Degree of Imagery (DOI)*

The defect of Hilbert Spectrum is that it's more discrete than the spectrums of other TFA methods, and it seems less meaningful to human eyes. Thus to obtain the feature, Degree of Imagery (DOI), the Hilbert Spectrum is processed by the following procedures.

First, a moving window is applied on the Hilbert Spectrum. For a time point t, its corresponding moving window is within the interval of [t+Δt], where Δt is the window length. For each moving window, the sub-spectrum is first added over the effective frequency range and then cumulated over the length of the window, which serves as a moving average filter over time.

$$E(t) = \int_{t-W}^{t} \int_{F_{lower\_bound}}^{F_{upper\_bound}} X(\tau, f) d\tau \qquad (5)$$

where $X(\tau, f)$ is the sub-spectrum of each moving window, W is the window length of the cumulation to calculate $E(t)$, and F_{upper_bound} and F_{lower_bound} is the effective frequency range.

DOI of each moving window is defined as follows.

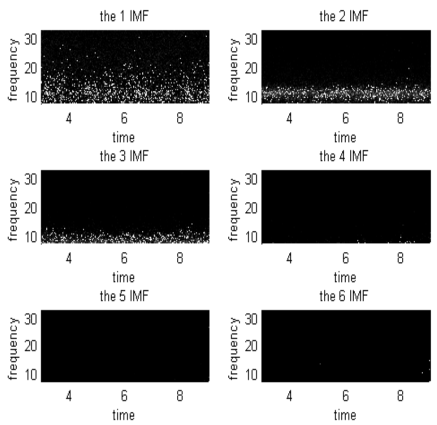


Figure 4. The top left figure shows the averaged Hilbert spectrum of the first IMFs of the 140 EEG signals recorded from 140 trials of motor imagery, and so on. The frequency range of each IMF is different from each other.

$$DOI(t) = \frac{Max(E_{C3}(t)) - Max(E_{C4}(t))}{Max(Max(E_{C3}(t)), Max(E_{C4}(t)))} \qquad (6)$$

Figure 5 would demonstrate how DOI improve the detection and classification of ERD effect. Figure 5(a) shows the original energy within the effective frequency range, X(t), of C3 and C4. It can be observed that the difference of energy of C3 and C4 isn't stable feature. While using the formula (5), Figure 5(b) shows the E(t) of C3 and C4 under a left hand imagery task. It's clearly that Ec3(t) is higher than Ec4(t), which corresponds to the contralateral ERD. In Figure 5(b), the physical meaning of DOI is obvious. It denotes ratio of the energy difference of the two electrodes to the energy of the non-ERD electrode, serving as the reference. If ERD is significant, the magnitude of DOI will become close to one. The value of DOI ranges from -1 to 1, where 1 denotes left hand imagery and vice versa.

Figure 5(c) shows the corresponding DOI of the same trial. The DOI ascends significantly from 0 to 1 and reaches 1 at around 6 sec. It's worth noticing that since the window length of the cumulation is 2.6 secs, the DOI at 6 sec contains the information from 3.4 sec to 6 sec, which corresponds to firing time of motor imagery

The advantage of DOI is its zero mean and its potential of denoising. It can be further shown that the MI of DOI is quite high in Figure 5(d), reaching the maximum near 0.7 bits, which leads DOI a good feature for the ERD application.

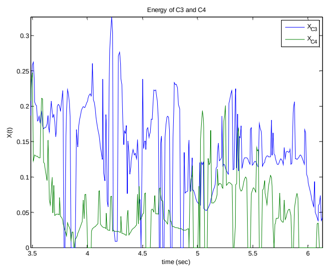
The length of the moving window is the factor to determine DOI, and it's obtained by calculating MI of DOI over the training data set with 10-folds cross-validation.

$$W = \frac{\sum w_i MI_i}{\sum MI_i} \qquad (7)$$

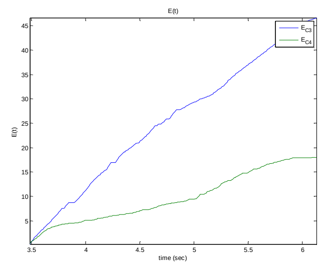
where w_i is the window length with the largest MI_i in each fold, and W is the window length to be determined.

With the determined window length, the MI of DOI versus time of the train data can be obtained. The data to train the classifier is then the DOI in the time interval when the MI is larger than 0.5. The classifier in the study is chosen to be Support Vector Machine (SVM), with either Linear kernel or Radial Basis Function kernel, and the parameters are too determined by 10-folds cross validation.

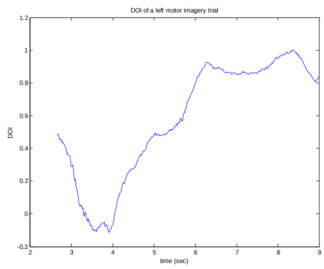
In the testing stage, if the mode is Offline Classification, the DOI of the test data is also truncated in the same time interval as in the training stage. On the other hand, if the mode is Online Simulation, the DOI will first be cumulated with the window size the same as in the training stage, i.e. for each time instant, there's a vector of features containing the DOI of the instant and before with the length of the window length. SVM would be performed over time and then for each time instant the output of SVM and classified result would be displayed.



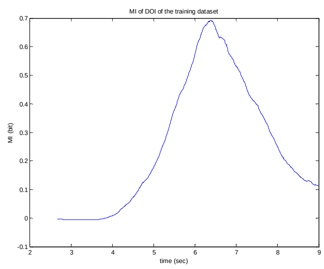
(a) Energy of C3 and C4



(b) E(t) of C3 and C4



(c) DOI of a left imagery trial



(d) MI of DOI of the training dataset

Figure 5. (a) shows the original energy within the effective frequency range, X(t), of C3 and C4. It can be observed that the difference of energy of C3 and C4 isn't stable feature. While using the formula (5), (b) shows the E(t) of C3 and C4 under a left hand imagery task. It's clearly that $E_{c3}(t)$ is higher than $E_{c4}(t)$, which corresponds to the contralateral ERD. (c) shows the corresponding DOI of the same trial. The DOI ascends significantly from 0 to 1 and reaches 1 at around 6 sec. It's worth noticing that since the window length of the cumulation is 2.6 secs, the DOI at 6 sec contains the information from 3.4 sec to 6 sec, which corresponds to firing time of motor imagery. (d) shows the MI of DOI versus time of the training dataset.

IV. Experimental Results

A. *The Previous Researches s*

In this section, the performance of the previous works using BCI competition 2003 dataset III will be examined. Since the intra-subject variance of EEG is large in BCI system, only the results with the same dataset would be comparable.

Originally, in the BCI competition 2003, the winners of dataset III were ranked by the highest MI value, using the test data in the dataset to perform the Online Simulation, classifying the trial over time,and the winner is S. Lemm, et al.[8], using complex Morlet Wavelet and probabilistic model, with highest MI 0.61 bits at 7.6 s and minimum error 10.7% at 6.8 s. Recently, many research apply their methods on this dataset to test the results. In this paragraph, the result of Offline Classification would be reviewed first and then the result of the Online Simulation.

Performing Offline Classification, one outcome per trial, N. Brodu [9] showed the comparative study of band-power extraction techniques. In this paper, the result of spectrogram is the best with accuracy 82.1%. Besides, Brodu [10] implemented the BCI system with multifractal cumulants and predictive complexity with accuracy 80.7%. B. Wang [11] investigates in the Gaussian Process Classifier with accuracy 86%. Y. Jiao [12] used AR model and combined SVM and GA, with accuracy 89.29 %. M.K. Hazrati [13] et al. used adaptive neural networks and gain the high accuracy up to 90%. X. Guo [14] adopted dynamic ICA mixing matrix with accuracy 87.14%.

As Online Simulation, H. Zhao [15] first used relative wavelet energy to discriminate and comes with highest MI 0.54 bits about 4.7s and later [16] using wavelet entropy with highest MI 0.62 bits about 4s. S. Rezaei et al. [17] compares the performance of different classifiers with highest MI around 0.5 bits and accuracy around 83.57%. S.M. Zhou et al. [18] used higher-order statistics with highest MI 0.64 bits and accuracy 90% around 6s.

The comparative table of the previous research is shown in Table 1.

B. *The Proposed Method*

In the dataset, there're total 280 trials and they are divided in two groups of dataset, training data and testing data, randomly and 140 trials each. The model of the proposed method is first trained using the training data, and the parameters are picked using 10-folds cross validation and the criterions addressed before. Then the performance of the model is validated on the testing data.

The experimental results of the proposed method are shown in Table 2 and Figure 6. Figure 6 shows how the MI and the accuracy change versus time of Online Simulation of different classifier and TFA methods, where the MI is calculated by the output of SVM. The result using HHT and linear kernel SVM comes with the highest MI 0.72 bits at around 7.1s, which is much higher than all the previous researches, and the result of HHT and radial basis kernel SVM is with accuracy 88.6 % and MI 0.69 bits. The results of Gabor Transform show the capability of generalization of the proposed method to other TFA methods.

TABLE I. THE RESULTS OF PREVIOUS WORKS

Offline Classification	**Accuracy (%)**
Bandpass Filter +Spectrogram+LDA [9]	82.1
Band Power + Multifractal Comulants + Predictive Complexity+LDA [10]	80.7
IIR Bandpass Filter Comulant Energy +Gaussian Process Classifier [11]	86.43
AR Model + GA-SVM [12]	89.29
Spectral of Windowed Segments +Adaptive Probabilistic Neural Network [13]	**90**
Dynamic ICA + Large Number Voting [14]	87.14
Online Simulation	***MI (bit)***
Complex Morlet Wavelet + 4-D Gaussian Model + Bayes Theorem [8]	0.61
Relative Wavelet Energy + SVM [15]	0.54
Wavelet Entropy + Band Powers + LDA [16]	0.62
AR + AAR+ Bayesian Network [17]	0.50
Higher Order Statistics +Neural Network [18]	**0.64**

TABLE II. THE EXPERIMENTAL RESULTS OF THE PROPOSED METHOD

Offline Classification	**Accuracy (%)**	
Gabor, SVM (linear)	87.1	
Gabor, SVM (rbf)	87.2	
HHT, SVM (linear)	90.0	
HHT, SVM (rbf)	**90.7**	
Online Simulation	***Maximum Accuracy (%)***	***Maximum MI (bit)***
Gabor, SVM (linear)	88.6	0.62
Gabor, SVM (rbf)	87.2	0.61
HHT, SVM (linear)	**90.0**	**0.72**
HHT, SVM (rbf)	88.6	0.69

V. DISCUSSION AND CONCLUSION

The experimental results shows the performance much better than others especially comparing MI. The importance of MI is that MI not only indicates the accuracy a classifier but also designates the strength of the output and how robust a system is to the noise. In the real application, the system with high classification accuracy does not always perform well because of the noise contaminating the signal especially in the EEG-based BCI system with low SNR, making the direct reading of the accuracy meaningless for losing the information of variance. Previously, most of the research focuses only on the outcomes of Offline Classification which makes the system may not be practical in use.

Though the proposed method using DOI and linear SVM is high in MI but the latency is a little too long. It's because there're two stages of cumulation, the cumulation to determine

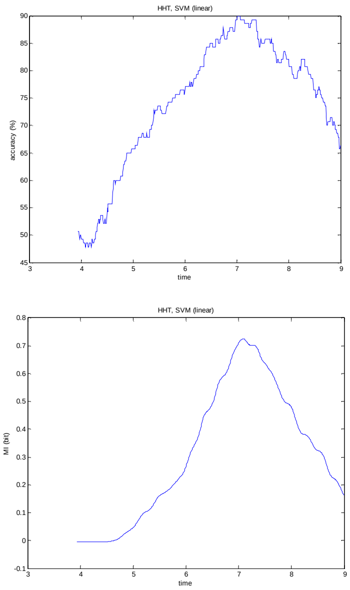


Figure 6. shows how the MI and the accuracy change versus time of Online Simulation of different classifier and TFA methods.

DOI and the SVM. DOI is gained by gathering the information of the spectral energy over time. SVM in the method uses a section of DOI to train and to classify. To resolve the problem, we can simply shorten the dimension of the feature in SVM, with a tradeoff of lower MI and accuracy, in the extreme case, leaving the SVM, only using the sign of raw DOI to classify. The classification of raw DOI can be carried out by the criterion as follows. If the sign of DOI is positive, the outcome is left hand motor imagery and vice versa. Since the maximum MI and the accuracy of raw DOI of the testing dataset in Figure 7 are around 89.2% and 0.71 bits around 6s, it limits the lower bound of the performance. Also, by using DOI only, no training of the classifier is needed, and the training of the window size in calculating DOI is much faster, usually in less than 5 seconds would largely shorten the training time of the system. Besides, the computational time of the classification is quite short. The system is implanted in Matlab using PC, with Core i5 750, and the computational time per second data is less than 0.02 cpu seconds thanks to the reduction in the iteration of HHT.

The experiment result of Gabor Transform replacing HHT shows the potential of the generalization of the notation of DOI to the class of Time-Frequency or Time-Scale Analysis. Different TFA method may lead to different representation and also some methods such as time-frequency reassignment can further improved the localization and clearness of the spectrum. It's believed that DOI can benefits from not only HHT but also a more sophisticated TFA method, such as EMD + Reassigned Spectrogram, however this is beyond the scope of this study.

It's interestingly that the best classification time interval determined by the method automatically by means of MI is from 3.2s to 7s, when MI of DOI is larger than 0.5 and taking the window length into consideration, which makes sense. It's possible that the subject is unable to focus and perform during

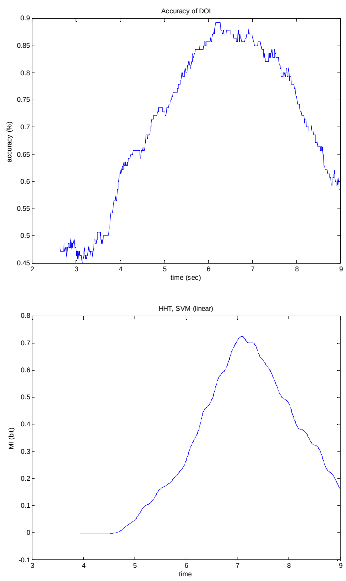


Figure 7. shows how the MI and the accuracy of the testing data change versus time of raw DOI.

the whole 6s motor imagery task. Further, it's possible of the usage of DOI to construct a motor imagery detection system.

In this paper, we have presented the novel feature, DOI, to improve the performance of motor imagery classification. The results carried out on the EEG data provided by the BCI 2003 competition has indicated the effectiveness of the proposed feature. The success of our method should be attributed to the use of the EMD process and the robust feature extraction method. The EMD process in HHT is an adaptive component extracting method, thus it's more suitable for EEG, and carries the ability of adaptive band filtering by selecting the corresponding IMFs; DOI is capable of extracting and amplifying the discriminating features from the spectrum of TFA methods such that the ERD is effectively detected.

REFERENCES

[1] L. Wang, G. Xu, J. Wang, S. Yang, and W. Yan, "Application of Hilbert-Huang Transform for the Study of Motor Imagery Tasks," presented at the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2008.

[2] P. Wei, Q. Li, and G. Li, "Classifying Motor Imagery EEG by Empirical Mode Decomposition Based on Spatial-Time-Frequency Joint Analysis Approach," presented at the International Conference on Future BioMedical Information Engineering, 2009.

[3] L. Yuan, B. Yang, S. Ma, and B. Cen, "Combination of Wavelet Packet Transform and Hilbert-Huang Transform for Recognition of Continuous EEG in BCIs," presented at the 2nd IEEE International Conference on Computer Science and Information Technology, 2009.

[4] 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," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903-995, 1998.

[5] A. Schlögl, *The electroencephalogram and the adaptive autoregressive model: Theory and applications*. Germany Shaker Verlag GmbH, 2000.

[6] A. Schlögl, "Outcome of the BCI-competition 2003 on the Graz data set," *url:* *http://www.bbci.de/competition/ii/*, 2003.

[7] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, pp. 112-114, 2004.

[8] S. Lemm, C. Schäfer, and G. Curio, "BCI Competition 2003—Data Set III: Probabilistic Modeling of Sensorimotor Rhythms for Classification of Imaginary Hand Movements," *IEEE Transactions on Biomedical Engineering*, vol. 51, June 2004.

[9] N. Brodu, F. Lotte, and A. Lecuyer, "Comparative Study of Band-Power Extraction Techniques for Motor Imagery," presented at the IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain, 2011.

[10] N. Brodu, F. Lotte, and A. Lécuyer, "Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity," *url:* *http://arXiv.org* 2010.

[11] B. Wang, F. Wan, P. U. Mak, P. I. Mak, and M. I. Vai, "EEG Signals Classification for Brain Computer Interfaces Based on Gaussian Process Classifier," presented at the 7th International Conference on Information, Communications and Signal Processing, 2009.

[12] Y. Jiao, X. Wu, and X. Guo, "Motor Imagery Classification Based on the Optimized SVM and BPNN by GA " presented at the International Conference on Intelligent Control and Information Processing, 2010.

[13] M. K. Hazrati and A. Erfanian, "An online EEG-based brain-computer interface for controlling hand grasp using an adaptive probabilistic neural network," *Medical Engineering & Physics*, vol. 32, pp. 730-739, 2010.

[14] X. Guo and X. Wu, "Motor Imagery EEG Classification based on Dynamic ICA Mixing Matrix " presented at the 4th International Conference on Bioinformatics and Biomedical Engineering, 2010.

[15] H. Zhao, X. Wang, and H. Wang, "Brain-computer interface design using relative wavelet energy," in *Chinese Control and Decision Conference*, 2008, pp. 3558-3561.

[16] H. Zhao, C. Liu, C. Li, and H. Wang, "Feature Extraction Using Wavelet Entropy and Band Powers in Brain-Computer Interface," presented at the 2nd International Conference on Signal Processing Systems, 2010.

[17] S. Rezaei, K. Tavakolian, A. M. Nasrabadi, and K. Setarehdan, "Different classification techniques considering brain computer interface applications," *Journal of Neural Engineering*, pp. 139-144, 2006.

[18] S. M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface," *Information Sciences*, vol. 178, pp. 1629-1640, 2008.