35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Multichannel EEG feature extraction based on Hilbert-Huang transform and extreme learning machine*

Min Han, *Senior Member, IEEE*, and Minghui Wang

Abstract— **EEG feature extraction is a difficult and important problem in EEG analysis which is significantly helpful in cerebral disease aided diagnosis, such as epilepsy. In this paper, a novel EEG feature extraction method is proposed, which obtains the phase series via Hilbert-Huang transform, and then the phase interaction information is extracted among all EEG channels by using neural networks. In the phase calculation, rather than the commonly used Hilbert transform and complex wavelet transform, the Hilbert-Huang transform is introduced, which is more suitable for nonlinear and nonstationary signal processing, decomposing and adaptive transforming. In the phase interaction information extraction, instead of a single synchronization index, an extreme learning machine is adopted, which can identify the phase interaction information via one-step prediction and taking the output weights as the features. Furthermore, the novel method is applied to epileptic seizure prediction. Simulations on Freiburg database show that the proposed method can extract the potential EEG characteristics well compared to other feature extraction methods, which are propitious to predict seizures more effectively.**

I. Introduction

Epilepsy is a kind of chronic brain dysfunction syndrome. About 25% of epilepsy patients cannot be treated sufficiently by any available therapy. They are at the risk of serious injuries, and an effective seizure prediction method is needed for them to provide warning time for safety-enhancing behavioral responses. Electroencephalograph (EEG) can reflect the physiological functions of the human brain and mental state. It is very helpful in monitoring of epilepsy.

EEG has been proven to be a kind of nonlinear and nonstationary time series [1]-[2]. Most of the researchers always focus their emphasis on the research of EEG feature extraction methods, which includes linear methods and nonlinear methods, or univariate measures and multivariate measures [3]. Linear methods mainly include spectral analysis [4], linear model analysis [5], and wavelet-based method [6]. Nonlinear methods mainly include correlation dimension [7], entropy-based approaches [8], and largest Lyapunov exponent [9]. All the above methods are mainly univariate measures. Multivariate measures include linear correlation analysis [10], and phase synchronization [11]-[13]. Reference [14] overviewed the most commonly used multivariate nonlinear features, and [15] presented the synchronization measures for analyzing EEG. A. Aarabi [3] pointed out that there is no clear superiority of the nonlinear measures over linear measures, whereas bivariate measures are generally more effective. In multivariate measures, synchronous analysis methods are the most widely studied.

The main idea of synchronous analysis methods is to analyze the phase synchronization between different channels, mainly concerning the phase obtaining and synchronization quantification [14]. In the aspect of phase calculation, Hilbert transform (HT) [11] and complex wavelet transform (CWT) [12] are most widely used; and in the aspect of phase synchronization, most researchers focused on the mean phase coherence (MPC) [11]-[14]. HT is simple, but it computes the instantaneous amplitude, frequency and phase of the signal within the mathematics framework in macro perspective. It is likely for negative frequency to occur, which does not make any sense. Meanwhile, when CWT is used, a proper wavelet needs to be selected, and also the transform result is not unique [16]. Therefore, in order to handle the problem of nonlinear and nonstationary signal analysis better, N.E. Huang proposed Hilbert-Huang transform (HHT) [17] in 1998, which can overcome the above-mentioned shortcomings of HT and CWT. HHT takes the smoothing processing at first, which can decompose original signal into several stationary ones without a special decomposition base, and the transform result is unique.

In addition to the transform methods for phase calculation, how to quantify the phase interaction among different channels is still a problem. Although MPC has a high computational efficiency, it quantifies the phase synchronization only between double channels from a single point of view, and unable to fully exploit the phase interaction relationships among all channels. Fortunately, neural networks can exploit the phase interaction information from the system identification point of view, analyzing the phase more comprehensively. Moreover, neural networks are nonparametric models, which are insensitive to the data distributions and characteristics, and appropriate for all kinds of patients [14]. Therefore, the neural networks are employed to build up the model in the feature extraction segment. In order to lower the computational cost and make it possible to use in online devices, extreme learning machine (ELM) [18] is chosen.

All the above considerations motivate the proposed method, which is a novel feature extraction method based on

*Research supported by the project of the National Nature Science Foundation of China under Grant 61074096.

M. Han is with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China (corresponding author to provide phone: +86-411-84708719; Fax: +86-411-84707847; E-mail: minhan@dlut.edu.cn).

M.H. Wang is with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China (E-mail: minghuiwang@mial.dlut.edu.cn).

HHT and ELM. HHT is used to calculate the phase of EEG, and ELM is adopted to quantify the interaction information among multichannels rather than double channels.

II. Multichannel EEG Feature Extraction based on HHT and ELM

In the phase analysis methods, there are two key points to focus on: phase calculation and phase interaction information extraction. According to the analysis in the Introduction part, a novel multichannel EEG feature extraction method based on HHT and ELM is proposed in this paper, which is named HHT-ELM for short. In general, HHT takes place of the commonly used HT and CWT, and ELM takes place of the phase synchronization indices (such as MPC) at the same time.

Fig. 1 shows the main structure of HHT-ELM. In Fig.1, the input of the whole structure is EEG with *d* channels, and they are transformed into phase series by HHT. Then, ELM is used to process the phase series. Through nonlinear mapping and one-step prediction training, the output weights **B** of ELM are obtained, which are taken as the EEG features we need.

A. HHT for Phase Calculation

HHT decomposes and transforms adaptively according to the data itself [17]. It consists of empirical mode decomposition (EMD) and HT [19]-[20]. The nature of EMD is time series smoothing processing, i.e., the different scales of fluctuations or trends of the upcoming complex signals are decomposed gradually [19]. Each scale is called as intrinsic mode function (IMF). For different signals, EMD has the adaptive decomposition ability and the decomposition result is unique.

Based on EMD, HHT can be explained as follows. For the given signal $x(t)$, EMD can decompose $x(t)$ into a group of IMFs, imf_i (i=1,2,…, n),

$$x(t)=\sum_{i=1}^{n} imf_i(t)+r(t)\,, \tag{1}$$

where n denotes the number of IMFs, $r(t)$ is called the residual function, representing the trend of signal $x(t)$. Then, applying HT to the IMF components, the following is obtained

$$Z(t)=imf(t)+jH[imf(t)]=a(t)\mathrm{e}^{j\int w(t)\mathrm{d}t}\,, \tag{2}$$

where,

$$a(t)=\sqrt{imf^2(t)+H^2[imf(t)]}\,,$$
$$\varphi(t)=\arctan\left(H[imf(t)]/imf(t)\right), \tag{3}$$
$$w(t)=\mathrm{d}\varphi(t)/\mathrm{d}\,t\,.$$

Through (2), the instantaneous amplitude $a(t)$, phase $\varphi(t)$, and angle frequency $w(t)$ of IMF can be obtained. $\varphi(t)$ calculated by (3) is the phase series we need.

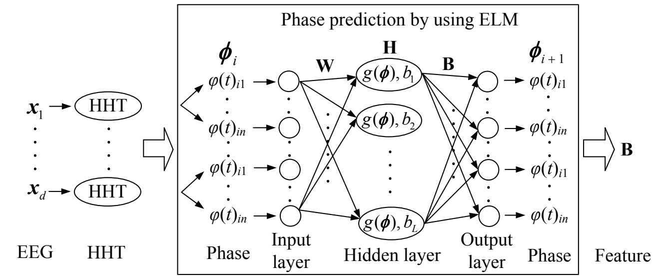

Figure 1. The structure of the multichannel EEG feature extraction based on HHT and ELM.

B. ELM for Phase Interaction Quantization

After calculating the phase, the phase interaction information needs to be extracted. ELM is utilized for imitating and identifying the phase interaction information among all channels with a low computation cost [18], instead of the commonly used phase synchronization indices [14].

Taking the phase interaction among different channels as a complicated system, then the quantization of that could be turned into a system identification problem. ELM is taken as the identifier, and the output weights are the system parameters to be identified. The inputs of the system are the current phases, and the outputs are the phases of the next time. Therefore, by means of one-step prediction of the phases, the phase interaction system can be identified, i.e., the phase interaction information among all channels can be quantified.

As shown in Fig.1, the input layer of ELM is phase $\boldsymbol{\phi}_i$, and the output layer of ELM is phase $\boldsymbol{\phi}_{i+1}$. Because the research of this paper is based on a moving-window analysis, therefore the feature extraction method acts on each time window. In each time window, the one-step prediction training procedure of ELM is used to fit the actual phase series. Then, the output weights **B** of ELM are obtained, which are taken as the phase interaction system parameters to be identified, i.e., the useful extracted EEG features of the corresponding time window. The features present all the information of the phase interaction among all channels.

ELM works for single-hidden layer feedforward networks [18], and it has been demonstrated to have impressive performance in regression and classification tasks due to high generalization ability and fast learning speed. Different from other general neural networks, ELM calculates the output weights using Moore-Penrose inverse rather than iterate, and generates the input weights and biases randomly rather than design them. Therefore, ELM can perform at extremely fast learning speed. In this paper, ELM is not only used in the feature extraction part, but also used as a classifier later.

III. Application in Epileptic Seizure Prediction

Epileptic seizures are usually characterized by an abnormal synchronized electric discharge of neurons involved in the epileptic process, implying that a method based on phase analysis should be adopted. Therefore, the proposed method HHT-ELM will be applied to the epileptic seizure prediction in this paper, and the interpretations are as follows.

Step 1: Preprocessing. In order to eliminate the influence of superimposed sinusoidal disturbance at the frequency of the ac power supply, a 50 Hz band-suppression filter is exploited.

Step 2: Feature extraction. The preprocessed EEG signals are passed through the novel feature extraction method HHT-ELM over time windows, producing a feature vector to be used for classification.

Step 3: Classification. ELM is also used to learn the mappings from the training set features into the patient's state: preictal or interictal.

Step 4: Alarm producing. From the classification results, not only the trend of patient's brain condition can be found, but also a chattering behavior can often be found. In order to avoid the chattering behavior which negatively affects the seizure prediction capability, "preictal density" *Den* in an observation window win_{os} is calculated

$$Den = N_{preictal} / (N_{preictal} + N_{interictal}), \tag{4}$$

where $N_{preictal}$ denotes the predicted preictal samples' number and $N_{interictal}$ denotes the predicted interictal sample's number, and a density threshold γ should be chosen. When *Den* is over γ, an alarm is produced, otherwise no alarm.

IV. Experiments and Results

To evaluate the proposed method, some simulations on the Freiburg EEG database (http://epilepsy.uni-freiburg.de) are carried out, which contains invasive EEG recordings of 21 patients suffering from medically intractable focal epilepsy. Because the EEG recordings come from cortex, so there is little volume conduction effect and we would not discuss whether the method will be affected by that or not.

A. Simulations

All the simulations were based on a 1.80 GHz 2-core CPU with 2.00 GB memory. The training and testing data sets have been generated for each patient separately. For the preictal data set, the first two seizures (or the first seizure for patient 8 and 13 with only two seizures totally) were used to produce the training set, and the remaining seizures were to test. By using the intervals of 10 s and overlapped by 50%, 37.6 minutes of data immediately preceding each seizure can produce 450 preictal training samples. For the interictal data set, the interictal training samples are also generated using the intervals of 10 s, randomly chosen from the interictal recordings of 24 h for a total of 900 interictal training samples; and the remaining record were to test.

The implementation of the proposed method also requires the choice of some design parameters. After some experimentation, we set the parameters as follows. The time window is set at 10 s, and overlapped by 50%. For HHT-ELM, the maximum number of IMFs is set at 3. Then according to the marginal spectrum of Hilbert-Huang spectrum we can see that, imf_1, imf_2, imf_3, and residual r mainly contain the component of 17~30Hz, 8~20Hz, 5~10Hz and 0~5Hz, respectively. They are nearly corresponding to EEG rhythms of β, α, θ and δ waves. The number of hidden neurons of ELM is empirically determined as 10, and the sigmoid function is chosen as the activation function. For ELM [21] for classification, the number of hidden neurons is set at 1000, and the activation function uses sigmoid functions. The observation window win_{os} is 1.5 min, and the density threshold γ is 0.7. In order to weaken the influence of the randomness caused by ELM, all the trials are repeated for 30 times, and the mean value of the thirty results is taken as the final result. For the comparison method, the parameters are as follows. We choose Gaussian complex wave as the base wave, i.e., Gaussian complex wave transform (GCWT), and the order of it is 4. The order of AR coefficient is 6, with the moving average parameters being 20, and estimated by Burg method.

B. Performance Evaluations

In order to illustrate the results clearly, the following evaluations are used: the sensitivity s_e (the percentage of seizures which have been predicted accurately), the false-positive rate *fpr* (the number of false alarms per hour in interictal EEG), the advance prediction time t_a (the difference between the seizure beginning time marked in the database and the alarm time determined by the prediction system), and the performance index *P* (shown in (5) [11]).

$$P = \sqrt{\left(\overline{s_e}^2 + \overline{s_p}^2\right)/2}, \tag{5}$$

where $\overline{s_e}$ denotes the mean sensitivity, and $\overline{s_p}$ denotes the specificity rate, which is defined as 1 minus the mean *fpr*. The larger *P*- the better the system.

C. Results

Table I and Table II show the comparison of different phase interaction quantization methods and the comparison of different transform methods, respectively. It can be seen from the two tables that, the mean t_a of them are all nearly 50 minutes. From Table I, compared with MPC, no matter transformed by HT or GCWT, not only s_e of ELM is higher, but also the mean *fpr* has an obvious advantage, i.e., *P* is much higher. From Table II, compared to HT and GCWT, not only s_e of HHT is higher, but also the mean *fpr* has an obvious advantage, i.e., *P* is much higher. In general, the results indicate that ELM is more effective than MPC for the quantization of the phase interaction, and HHT is more suitable for nonlinear and nonstrationary signal analysis.

In order to illustrate that the proposed seizure prediction system is more effective, the proposed method (marked as "HHT-ELM-ELM") is compared with two other popular methods: the classical phase synchronization method [11] (marked as "Method 1") and the method based on AR model [5] (marked as "Method 2"). Method 1 uses an automated technique for detecting decreased synchronization, finding of a preictal drop in synchronization, and distinguishes the preictal state from the interictal interval. Method 2 uses AR coefficients as EEG features, and takes ELM as the classifier. The results are compared in Table III.

TABLE I. RESULTS COMPARISON OF PHASE INTERACTION QUANTIZATION METHODS MPC AND ELM

Feature extraction method	HT-MPC	HT-ELM	GCWT-MPC	GCWT-ELM
Mean s_e (%)	70.7	**77.2**	72.8	**73.9**
Highest s_e (%)	87.0	**89.1**	84.8	**84.8**
Mean t_a (min)	49.7	48.0	56.0	53.6
Mean *fpr* (h^{-1})	0.27	**0.24**	0.22	**0.14**
P	0.72	**0.77**	0.76	**0.80**

TABLE II. RESULTS COMPARISON OF TRANSFORM METHODS HT, GCWT AND HHT

Feature extraction method	HT-ELM	GCWT-ELM	HHT-ELM
Mean s_e (%)	77.2	73.9	**84.8**
Highest s_e (%)	89.1	84.8	**95.7**
Mean t_a (min)	48.0	53.6	52.3
Mean *fpr* (h^{-1})	0.24	0.14	**0.08**
P	0.77	0.80	**0.89**

TABLE III. RESULTS COMPARISON OF DIFFERENT METHODS

Items	Method 1	Method 2	HHT-ELM-ELM
Mean s_e (%)	71.7	81.5	**84.8**
Highest s_e (%)	71.7	87.0	**95.7**
Mean t_a (min)	40.2	61.8	52.3
Mean *fpr* (h^{-1})	0.35	0.11	**0.08**
P	0.68	0.85	**0.89**

In Table III, the mean t_a of the three methods are all over 40 minutes. The mean sensitivity of our method HHT-ELM-ELM is 84.8%, the mean false-positive rate is 0.08 h^{-1}, and the performance index is 0.89. Compared to Method 1, the proposed system has the obviously better performance, also shows that the introduction of machine learning can improve the sensitivity and specificity at the same time. Compared to Method 2, both of the mean s_e and highest s_e of the proposed method are higher, and at the same time, the mean *fpr* is lower. *P* of the proposed method is the highest one, indicating that our method performs the best.

V. CONCLUSIONS

A new multichannel EEG feature extraction method has been proposed, which is consisting of HHT and ELM. HHT is utilized instead of commonly used HT and CWT, and ELM is adopted to extract the phase interaction information rather than MPC. In order to evaluate the performance of the proposed method, it was applied to the epileptic seizure prediction system, and careful comparison experiments on the Freiburg database were carried out. The results clearly indicate that the proposed method offers a better balance of sensitivity and false-positive rate. HHT is more suitable for nonlinear and nonstationary signal analysis, and HHT-ELM extracts the features representing the phase interaction information among all channels well.

REFERENCES

[1] X. Y. Wang and C. Luo, "Researches on chaos phenomenon of EEG dynamics model," *Applied Mathematics and Computation*, vol. 183, no. 1, pp. 30-41, Dec. 2006.

[2] B. He, L. Yang, C. Wilke, and H. Yuan, "Electrophysiological imaging of brain activity and connectivity—challenges and opportunities," *IEEE Trans. on Biomed. Eng.*, vol. 58, no. 7, pp. 1918-1931, Jul. 2011.

[3] A. Aarabi, R. Fazel-Rezai, and Y. Aghakhani, "EEG seizure prediction: measures and challenges," in *31th Annu. Int. Con. of the IEEE Eng. Med. Biol. Soc.*, Minneapolis, MN, 2009, pp. 1864-1867.

[4] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Analysis of epileptic EEG signals using higher order spectra," *J. of Med. Eng. and Technol.*, vol. 33, no. 1, pp. 42-50, Jan. 2009.

[5] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, et al, "Real-time epileptic seizure prediction using AR models and support vector machines," *IEEE Trans. on Biomed. Eng.*, vol. 57, no. 5, pp. 1124-1132, May 2010.

[6] A. S. Zandi, M. Javidan, G. A. Dumont, and R. Tafreshi, "Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform," *IEEE Trans. on Biomed. Eng.*, vol. 57, no. 7, pp. 1639-1651, Jul. 2010.

[7] A. F. Rabbi, A. Aarabi, and R. Fazel-Rezai, "Fuzzy rule-based seizure prediction based on correlation dimension changes in intracranial EEG," in *32nd Annu. Int. Conf. of the IEEE Eng. in Med. and Biol. Soc.*, Buenos Aires, Argentina, 2010, pp. 3301-3304.

[8] H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, no. 2, Part 1, pp. 2027-2036, Mar. 2009.

[9] S. Sabesan, N. Chakravarthy, K. Tsakalis, P. Pardalos, and L. Iasemidis, "Measuring resetting of brain dynamics at epileptic seizures: application of global optimization and spatial synchronization techniques," *Journal of Combinatorial Optimization*, vol. 17, no. 1, pp. 74-97, Jan. 2009.

[10] M. Tito, M. Cabrerizo, M. Ayala, A. Barreto, L. Miller, *et al*, "Classification of electroencephalographic seizure recordings into ictal and interictal files using correlation sum," *Computers in Biology and Medicine*, vol. 39, no. 7, pp. 604-614, Jul. 2009.

[11] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, *et al*, "Epileptic seizures are preceded by a decrease in synchronization," *Epilepsy Research*, vol. 53, no. 3, pp. 173-185, Mar. 2003.

[12] L. Wang, C. Wang, F. Fu, X. Yu, H. Guo, *et al*, "Temporal lobe seizure prediction based on a complex Gaussian wavelet," *Clinical Neurophysiology*, vol. 122, no. 4, pp. 656-663, Apr. 2011.

[13] M. Han, M. H. Wang, X. J. Hong, and J. Han, "Epileptic seizure prediction based on probabilistic discriminative extreme leaning machine," *Chinese Journal of Biomedical Engineering*, vol. 31, no. 2, pp. 175-183, Apr. 2012.

[14] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, no. 1-2, pp. 1-37, Sep.-Oct. 2005.

[15] V. Sakkalis, C. Doru Giurcaneanu, P. Xanthopoulos, M. E. Zervakis, V. Tsiaras, *et al*, "Assessment of linear and nonlinear synchronization measures for analyzing EEG in a mild epileptic paradigm," *IEEE Trans. on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 433-441, Jul. 2009.

[16] G. Jager, R. Koch, A. Kunoth, *et al*, "Fast empirical mode decompositions of multivariate data based on adaptive spline-wavelets and a generalization of the Hilbert-Huang-transformation (HHT) to arbitrary space dimensions," *Advances in Adaptive Data Analysis*, vol. 2, no. 3, pp. 337-358, Jul. 2010.

[17] N. E. Huang, and Z. Shen, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society 454(A)*, 1998, pp. 903-995.

[18] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *Int. J. of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107-122, Jun. 2011.

[19] G. Rilling, "Empirical Mode Decomposition," Available: http://perso.ens-lyon.fr/patrick.flandrin/emd.html

[20] D. Chen, D. Li, M. Z. Xiong, H. Bao, and X. L. Li, "GPGPU-aided ensemble empirical-mode decomposition for EEG analysis during anesthesia," *IEEE Trans. on Information Technology in Biomedicine*, vol. 14, no. 6, pp. 1417-1427, Nov. 2010.

[21] G. B. Huang, "MATLAB Codes of ELM Algorithm," Available: http://www.ntu.edu.sg/home/egbhuang/ELM_Codes.htm