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

Differential Entropy Feature for EEG-based Vigilance Estimation

Li-Chen Shi, Ying-Ying Jiao and Bao-Liang Lu* *Senior Member*, *IEEE*

Abstract— **This paper proposes a novel feature called differential entropy for EEG-based vigilance estimation. By mathematical derivation, we find an interesting relationship between the proposed differential entropy and the existing logarithm energy spectrum. We present a physical interpretation of the logarithm energy spectrum which is widely used in EEG signal analysis. To evaluate the performance of the proposed differential entropy feature for vigilance estimation, we compare it with four existing features on an EEG data set of twenty-three subjects. All of the features are projected to the same dimension by principal component analysis algorithm. Experiment results show that differential entropy is the most accurate and stable EEG feature to reflect the vigilance changes.**

I. INTRODUCTION

Vigilance means the degree of wakefulness or responsiveness to stimuli in medical definition. In cognitive neuroscience, the term of vigilance is used to describe the ability to sustain attention to a task for a long time. Some scientists use vigilance to imply the degree of arousal on the sleep-wake axis. Therefore, vigilance analysis can be used to index the drowsiness of drivers [1], [2], [3].

Up to now, various physiological measures are studied for vigilance analysis, such as eye movement [4] and autonomic nervous system activity. Among them, electroencephalogram (EEG) is the most commonly studied signal, and has been proved very effective [5], [6], [7], [8].

EEG as a time series signal, the potential change information of original form usually can not be directly used for vigilance estimation in non ERP experiments. Therefore, feature extraction is required. The extracted features should describe the differences of EEG patterns at different states of vigilance. Various EEG features can be classified into four categories: time series feature [9], spectral feature [5], spatial synchronization feature [10], and complexity measure feature [11].

Recently, a number of different entropy estimators have been applied to quantify the complexity of EEG signals [12], [13]. Entropy is a thermodynamic quantity describing the amount of disorder in the system. From an information theory perspective, the concept of entropy is generalized as the amount of information stored in a more general probability distribution. The concept of entropy has been successfully applied to the analysis of EEG signals.

L. C. Shi, Y. Y. Jiao, and B. L. Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University and MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.
* Corresponding author (bllu@sjtu.edu.cn)

In this study, differential entropy is proposed to characterize the level of vigilance, as a complexity measure for continuous EEG signals. Although energy spectrum and the logarithm energy spectrum have been widely used in EEG-based vigilance analysis, the specific meaning of logarithm power spectrum is not clear. Previous studies of estimating driving performance have confirmed that the EEG spectral amplitudes correlate with the wake-sleep transition more linearly in logarithmic scale than in linear scale [6], [14]. In this paper, we show that the proposed differential entropy, which is equivalent to logarithm energy spectrum, is superior to energy spectrum [5] and even other features such as auto-regressive (AR) parameters [15], fractal dimension [16], and sample entropy [17]. A detail performance comparison among these five different features is given by analyzing an EEG data set of twenty-three subjects in performing a monotonous visual task.

This paper is organized as follows. In section II, the detail experiment setup of a monotonous visual task is described. In section III, the differential entropy for characterizing EEG signals is presented and discussed in detail. In section IV, data analysis procedure is introduced and the experimental results are discussed. Section V concludes this study.

II. MATERIALS

A. Procedure and Subjects

A monotonous visual task was adopted here [8]. Subjects sat in a comfortable chair, about two feet away from the LCD. Four colors of traffic signs were presented in the LCD randomly by the NeuroScan Stim2 software. For each color, there were more than 40 different traffic signs. Each trial was $5.5s \sim 7.5s$ long, including $5s \sim 7s$ black screen and 0.5s traffic signs presented. The subjects were asked to recognize the sign color, and pressed the correct button on the response pad as soon as possible. There were 4 buttons on the response pad corresponding to the 4 different colors of traffic signs.

A total of twenty-three healthy subjects aged from 19 to 28 participated in this experiment. They were required to abstain from alcohol and caffeine one day before and during the experiment. After training, each subject took at least two sessions on different days, which were carried out in a small sound proof room with normal illumination. Each session lasted for more than 1 hour during $13:00 \sim 15:00$ after lunch.

B. Data Collection

For each session, the visual stimulus sequence and subject's response sequence were recorded by the NeuroScan Scan software. At the same time, a total of 62 EEG channels were recorded by the NeuroScan system at sampling rate 500Hz and then re-sampled down to 100Hz for the simplicity of data processing. In this study, due to the algorithm for finding the key brain area in our previous work, only 6 EEG channels (PO3, POz, PO4, O1, O2, Oz) around the occipital region were used for vigilance estimation to improve the feasibility of EEG-based vigilance estimation system.

C. Vigilance Measurement

In the sustained visual task, when becoming sleepy, the subject were more likely to press the wrong button or miss some responses. To evaluate the subject's vigilance level, the local error rate $e(t)$ of a subject's performance was used as the reference of vigilance levels, which was defined as the rate that the subject made a false response (including lapse) within a time window with a constant width [18].

III. DIFFERENTIAL ENTROPY FEATURE

A. Differential Entropy

Differential entropy is used to measure the complexity of a continuous random variable and is the entropy of continuous random variable. Differential entropy is also related to minimum description length. Its calculation formula can be expressed as,

$$h(X) = -\int_X f(x)\log(f(x))dx \tag{1}$$

where X is a random variable, $f(x)$ is the probability density function of X. For the time series X obeying the Gauss distribution $N(\mu, \sigma^2)$, its differential entropy can be defined as,

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}})dx \\ &= \frac{1}{2}\log(2\pi e\sigma^2) \end{aligned} \tag{2}$$

Although the original EEG signals do not follow a certain fixed distribution, it can be found that EEG signals are subject to Gaussian distribution nearly in a series of sub-bands after band-pass filtering from 2Hz to 44Hz by every 2Hz step. The distribution of EEG signals in such sub-bands can be verified by Kolmogorov-Smirnov test method, which is based on the cumulative distribution function, and tests whether a cumulative distribution is in compliance with certain theory distribution or whether a significant difference exists between two experience distributions. If the test result h is 1 under the significant level α, it means the hypothesis H is rejected that the sample series conforms to the specified gaussian distribution under this significant level. Otherwise, if the result h is 0, it means the hypothesis H is established.

In order to test whether the EEG signals after band-bass filtering follow Gaussian distribution, 2000 EEG data segments of 2 second length in every sub-band signal from 23 subjects' near occipital brain areas are randomly selected. Then, the Kolmogorov-Smirnov test method is applied to all 2000 EEG data segments to test whether each sub-band signal is subject to Gaussian distribution, under the significant level α set to 0.05. It can prove that the probability of sub-band signals meeting Gaussian distribution hypothesis is more than 90 percent.

Therefore, in a fixed frequency band i, the differential entropy is defined as

$$h_i(X) = \frac{1}{2}\log(2\pi e\sigma_i^2) \tag{3}$$

where h_i and σ_i^2 denote the differential entropy of the corresponding EEG signal in frequency band i and the signal variance, respectively.

B. Differential Entropy versus logarithm Energy Spectrum

The relationship between the logarithm energy spectrum and the differential entropy will be interpreted by the following mathematical derivation in each frequency band i of EEG signals.

From Eq. (3), we can see that π, e are constants, so only signal variance σ_i^2 is needed to know in order to calculate differential entropy $h_i(X)$ of Gaussian signal sequence $\{X\}$. For a frequency band of the EEG signals, as a result of zero mean (DC component is filtered), variance can be estimated by

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=0}^{N-1} x_i^2 \tag{4}$$

Equation (4) shows that the variance estimation of signal sequence X is just its average energy. What is more, average energy is related to energy spectrum due to Parseval's theorem. It can be written as Eq. (5) for the discrete Fourier transform,

$$\sum_{n=0}^{N-1} |x_i^2| = \frac{1}{N}\sum_{k=0}^{N-1} |X_k^2| = P_i \tag{5}$$

where $\{X_k\}$ is the discrete Fourier transform coefficients of signal sequence $\{x_i\}$. P_i can be regarded as energy spectrum and is equivalent to the value of signal variance σ_i^2 multiplying a constant coefficient N (N is the length of the fixed time window).

Therefore, after band-pass filtering to a specific frequency band i, its variance can be estimated by N/P_i. From Eqs. (3), (4) and (5), the relationship between the logarithm energy spectrum and the differential entropy can be expressed as,

$$\begin{aligned} h_i(X) &= \frac{1}{2}\log(2\pi e\sigma_i^2) = \frac{1}{2}\log(N\sigma_i^2) + \frac{1}{2}\log(\frac{2\pi e}{N}) \\ &= \frac{1}{2}\log(P_i) + \frac{1}{2}\log(\frac{2\pi e}{N}) \end{aligned} \tag{6}$$

It is thus clear that, for a fixed length EEG sequence, the estimation of differential entropy is equivalent to the logarithm energy spectrum in a certain frequency band. Researchers did often use energy spectrum and logarithm energy spectrum. Because the low frequency energy is often higher than the high frequency energy in EEG, so after the logarithm of energy, the ability of discriminating EEG pattern can be balanced between high and low frequency energy. Now, the physical meaning of the logarithm energy spectrum or energy spectrum can be explained from the differential entropy angle.

IV. Data Analysis Procedure

The procedure of data analysis for estimating the level of vigilance consisted of the following seven steps.

1) De-noising: EEG data were first preprocessed by using a simple band-pass filter between 0.1 and 100Hz and a cut-off frequency of 50Hz to remove noises with NeuroScan software.

2) Feature extraction: For each subject, each channel of all 6 channels was first filtered in many frequency bands by every 2Hz, from 1Hz to 44 Hz, and a 200-point Hanning window with non overlap was used to divide each frequency band signal to many window data of 2 second length. The differential entropy feature was calculated over each window data by FFT. Thus, the time series features for each session consisted of 6-channel EEG differential entropies estimated across 22 frequency bands.

3) Feature smoothing: The moving average filter with the window length 2 minutes was used to smooth the features, which can remove the influences of artifacts to a degree after feature extraction.

4) Feature dimension reduction: A standard PCA algorithm was used to obtain the directions of largest variance for each session. The differential entropy was reduced to 10 dimensions, and was used as inputs to train the individual regression model.

5) Feature selection: Feature selection based linear correlation coefficient was used. This method first calculated the linear correlation coefficient between train set and vigilance level curve, and then the features corresponding to larger correlation coefficient were selected. The optimal number of features was determined by cross validation.

6) Regression model: A support vector regression (SVR) model was used to estimate the time course of the driving performance. For SVR model, LIBSVM package was used and radial basis function (RBF) was selected. The range of the penalty factor C and the parameter γ of RBF were set to [0,1024] and [0.1,2], respectively. All of the points of (C,γ) were tried to find the best test result.

7) Vigilance level evaluation: The performance of the differential entropy was evaluated by the index of root mean square error (RMSE) between the estimated error rate and the true error rate curve(the local error rate $e(t)$).

V. Results and Discussion

The experimental results over 23 subjects' test sets are depicted in Fig. 1. The performance of the differential entropy was evaluated by the values of RMSE, and was compared to other four features, which were energy spectrum, auto-regressive parameters, sample entropy and fractal dimension. The RMSE value indicated the mean of each subject performance. The smaller the value of RMSE, the more precise the vigilance estimation. Processing procedure of other four features was similar to that of the differential entropy explained in section IV. Local error rate was taken as the actual performance of a training session with respect to a subject. The SVR model was trained with a training session and tested against a separate test session with respect to the same subject.

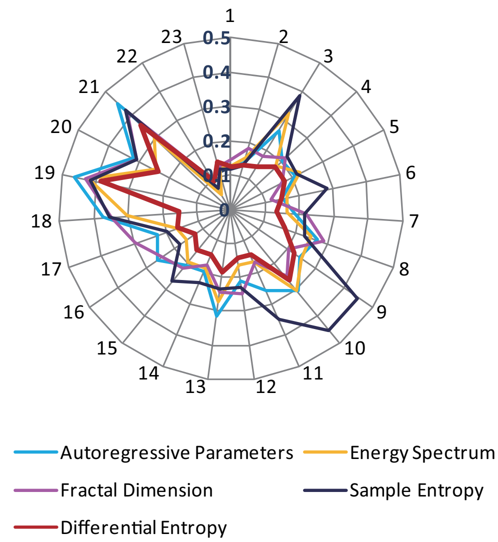

Fig. 1. Performance comparison among five features in vigilance estimation. Here, the label of subject is from 1 to 23, and the range of RMSE value is in [0.1,0.5].

As can be seen from Fig. 1, the RMSE values corresponding to the differential entropy, marked by the red line, were generally lower than those corresponding to other four features. According to Fig. 1, Table I presented the mean and variance of all five features over the data set of twenty-three subjects. After average of these results of RMSE values, the performance gap can be found among these five features. As can be seen from Table I, the mean RMSE value of the differential entropy was the smallest and it is 0.179, and its standard deviation was much smaller. The second small value corresponds to the energy spectrum (ES), and was followed by the fractal dimension (FD) and the AR parameters (ARP) in turn. The last one was the sample entropy (SE) feature.

TABLE I
The Mean and Variance of RMSE values of five features over the data set of twenty-three subjects.

	Feature Extraction methods				
	SE	ARP	FD	ES	DE
Average±sd	0.264 ±0.111	0.243 ±0.099	0.231 ±0.087	0.213 ±0.082	**0.179 ±0.072**

For the five features mentioned above, Table I focused on the analysis of ability of characterizing vigilance EEG, while Table II focus on the analysis of time complexity and robustness (anti-noise ability). Here, K was the number of electrodes, N was the number of EEG sequence samples, p was the order of AR model, and m was the length of sub sequence in the procedure of calculating sample entropy. As can been seen from Table II, the time complexity of sample entropy was minimum, while no much difference exists at time complexity among the other four features.

TABLE II
COMPARISON OF FIVE DIFFERENT FEATURE EXTRACTION ALGORITHMS IN TIME COMPLEXITY AND ROBUSTNESS.

Feature Extraction Algorithms	**Performance Index**	
	Time Complexity	Robustness
AR Parameters	$O(KNp + Kp^3)$	Medium
Energy Spectrum	$O(KN \log N)$	High
Fractal Dimension	$O(KN)$	High
Sample Entropy	$O(Km(N - m)^2)$	Low
Differential Entropy	$O(KN \log N)$	High

In terms of robustness, energy spectrum, differential entropy and fractal dimension had the highest robustness, but sample entropy was the worst. The main reasons were as follows: the energy spectrum and differential entropy were mainly based on FFT transform, which was linear. In the procedure of calculating energy spectrum and differential entropy, if the noise in EEG was limited to only a few frequency band, the rest frequency band will not be affected. The calculation of fractal dimension was subject only to linear interference. Because in the process of calculating sample entropy, the noise of interference will affect the overall and this interference was nonlinear, thus a greater impact would be generated on the results. If the noise in EEG was white, it will not affect the calculation of AR parameters, otherwise the nonlinear disturbance would generate, thus serious impact on the solution result will be caused.

From section III, we knew by estimating the Gauss parameters, the relationship between the differential entropy and the log power spectrum can be obtained. Consequently, the logarithm energy spectrum was equivalent to the differential entropy and also had better performance than other four features above. Furthermore, we can explain the logarithm energy spectrum from the differential entropy perspective.

According to the information theory, entropy can be considered to be a measure of the degree of 'disorder' of the system. Therefore, the differential entropy changes in many different frequency band can describe the degree of disorder of EEG data in vigilance experiment. Frequency band signal with a broad flat probability distribution will have high entropy, and frequency band signal with a narrow peaked distribution will have low entropy. Both the experiment results and analysis mentioned above indicated that the proposed differential entropy had higher capability to reflect the changes of vigilance. Meanwhile, entropy measures demonstrated promising performance in analyzing EEG signals related to vigilance.

VI. CONCLUSION

This paper presented a physical interpretation of logarithmic form of spectrum feature in EEG signal analysis and defined a new feature: differential entropy. A systematic comparative study on an EEG data set of twenty-three subjects in performing a monotonous visual task demonstrated that the proposed differential entropy feature was superior to four existing features to represent the changes of vigilance and raised the estimation accuracy by 5.9% in average.

VII. ACKNOWLEDGEMENT

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61272248) and the National Basic Research Program of China (Grant No. 2009CB320901; Grant No. 2013CB329401), and the European Union Seventh Framework Program (Grant No.247619).

REFERENCES

[1] B. Oken, M. Salinsky, and S. Elsas, "Vigilance, alertness, or sustained attention: physiological basis and measurement," *Clinical Neurophysiology*, vol. 117, no. 9, pp. 1885–1901, 2006.

[2] C. Cajochen, J. Zeitzer, C. Czeisler, and D. Dijk, "Dose-response relationship for light intensity and ocular and electroencephalographic correlates of human alertness," *Behavioural Brain Research*, vol. 115, no. 1, pp. 75–83, 2000.

[3] M. Weinger, "Vigilance, boredom, and sleepiness," *Journal of Clinical Monitoring and Computing*, vol. 15, no. 7, pp. 549–552, 1999.

[4] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imaging*, vol. 8, no. 5, pp. 357–377, 2002.

[5] T. Jung, S. Makeig, M. Stensmo, and T. Sejnowski, "Estimating alertness from the eeg power spectrum," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 1, pp. 60–69, 1997.

[6] C. Lin, R. Wu, T. Jung, S. Liang, and T. Huang, "Estimating driving performance based on eeg spectrum analysis," *Journal on Applied Signal Processing*, vol. 19, no. 3, pp. 3165–3174, 2005.

[7] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–130, 2011.

[8] L. Shi and B. Lu, "Eeg-based vigilance estimation using extreme learning machines," *Neurocomputing*, vol. 102, pp. 135–143, 2013.

[9] C. Anderson, E. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 3, pp. 277–286, 1998.

[10] F. Mormann, K. Lehnertz, P. David, and C. E Elger, "Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients," *Physica D: Nonlinear Phenomena*, vol. 144, no. 3, pp. 358–369, 2000.

[11] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[12] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic eeg detection using artificial neural networks," *IEEE Trans. Infor. Technology in Biomedicine*, vol. 11, no. 3, pp. 288–295, 2007.

[13] D. Abásolo, R. Hornero, P. Espino, D. Alvarez, and J. Poza, "Entropy analysis of the eeg background activity in alzheimer's disease patients," *Physiological Measurement*, vol. 27, no. 3, p. 241, 2006.

[14] A. Destexhe, D. Contreras, and M. Steriade, "Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states," *The Journal of Neuroscience*, vol. 19, no. 11, pp. 4595–4608, 1999.

[15] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabi. Eng.*, vol. 6, no. 3, pp. 316–325, 1998.

[16] H. Preißl, W. Lutzenberger, F. Pulvermüller, and N. Birbaumer, "Fractal dimensions of short eeg time series in humans," *Neuroscience Letters*, vol. 225, no. 2, pp. 77–80, 1997.

[17] H. Al-Angari and A. Sahakian, "Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1900–1904, 2007.

[18] S. Makeig and M. Inlow., "Lapse in alertness: coherence of fluctuations in performance and eeg spectrum," *NeuroImage*, vol. 86, no. 1, pp. 23–25, 1993.