Phase Lag Index Histogram Features for Identifying Epileptic Seizure EEG Signals

Masashi Nishiyama*, Yoshio Iwai*, Masayoshi Oguri† and Yoshihiro Maegaki‡
*Graduate School of Sustainability Science, Tottori University, Japan
†Faculty of Health Sciences, Kagawa Prefectural University Of Health Sciences, Japan
‡Faculty of Medicine, Tottori University, Japan

***Abstract*—Recent medical analytical studies have focused on the phase lag index (PLI) to analyze the functional connectivity of the brain. This paper assumes that the PLI contains informative characteristics for automatically identifying electroencephalogram (EEG) signals of epilepsy patients. We propose a method for distinguishing between epileptic seizure and non-seizure EEG signals using PLI histograms acquired from the signals of a short period, randomly sampled from longer recordings, in different brain regions. We demonstrate the ability of the method to identify epileptic seizures by experiments on the publicly available CHB-MIT Scalp EEG database.**

***Index Terms*—Electroencephalogram signals, epileptic seizures, histogram features, phase lag index.**

I. Introduction

Epilepsy is a neurological disorder in which severe seizures are sometimes caused by the abnormal activity of brain functions. The prevalence of epilepsy is quite high, at about one person in 100. For this reason, a doctor may diagnose many epilepsy patients every day, using EEG monitoring systems. In general, the EEG signals are repeatedly recorded across several days. In rare cases, the signals are recorded continuously, 24 hours/day, during several days. In practice, a doctor directly reads long-duration EEG recordings to find seizure characteristics. It is said that reading long EEG recordings takes substantial additional time—about one third of the recording time—and so it is a heavy burden for a doctor. There is therefore a need for a support system for primary screening, to distinguish between seizure EEG signals and non-seizure EEG signals and thereby assist a doctor in diagnosing epilepsy. For such a system, it is essential to design a method for accurately identifying seizure EEG signals.

To design a method for identifying epileptic seizure EEG signals, we focus on the functional connectivity of the brain. As described in [1], functional connectivity refers to the degree of correlation in activity between spatially distant brain regions. In the medical field, it has been reported that epilepsy impairs the functional connectivity of the brain [2], [3]. Recently, medical analytical studies [4], [5] have reported that the use of the phase lag index (PLI) is effective for analyzing the functional connectivity of the brain. However, these studies did not consider epileptic seizure EEG signals. Furthermore, there is no method for automatically distinguishing between epileptic seizure EEG signals and non-seizure EEG signals using the characteristics of the PLI.

We propose a method for distinguishing between epileptic seizure and non-seizure EEG signals using PLI histograms. Our method extracts PLI histogram features by applying a bandpass filter and computing the PLI values of EEG signals during randomly sampled short periods. Experimental results, using the publicly available CHB-MIT Scalp EEG database, show that our method using PLI histogram features has the ability to identify epileptic seizure EEG signals.

II. Phase Lag Index

Before explaining the detail of our histogram features, we explain PLI [6] and how it is used for analyzing the functional connectivity of the brain. Phase synchronization indices such as coherence and phase-locking value are widely used to represent the degree of connectivity. However, these indices are erroneously increased by volume conduction[1]. This problem is solved by PLI, which is another phase synchronization index. Stam et al. define the PLI value p between two signals as

$$p = \left| \frac{1}{T} \sum_{t=1}^{T} \mathrm{sgn}\left(\Delta\phi(t)\right) \right|, \tag{1}$$

where $\Delta\phi(t) = \phi^m(t) - \phi^n(t)$ is the phase difference between the two signals, $\mathrm{sgn}(\Delta\phi(t))$ is the signum function of $\Delta\phi(t)$, $t \in \{1, \cdots, T\}$ is the (discrete) time, T is the sampling length of the signals, and $\phi^m(t)$ and $\phi^n(t)$ are the phase components of the two signals. The range of $\Delta\phi(t)$ is defined as $-\pi < \Delta\phi(t) \leq \pi$. $\Delta\phi(t)$ is computed from two complex signals converted using Hilbert transform, as we describe in detail in (8). The range of p is $0 \leq p \leq 1$. When $p = 0$, phase synchronization does not appear or phase synchronization appears only when $\Delta\phi(t)$ is close to 0 or $\pm\pi$. When $p = 1$, phase synchronization perfectly appears except when $\Delta\phi(t)$ is close to 0 or $\pm\pi$. In general, p takes a large value when phase synchronization between the two signals strongly appears.

III. Our PLI Histogram Features

A. Overview

Here we outline our PLI histogram feature for distinguishing between epileptic seizure and non-seizure EEG signals. To design the feature, we used knowledge of the diagnosis

[1]Volume conduction: Suppose that neurons and synapses change their electrical potential because of activity. The potential change diffuses and reaches the electrodes on the scalp through the cerebrospinal fluid and skull.

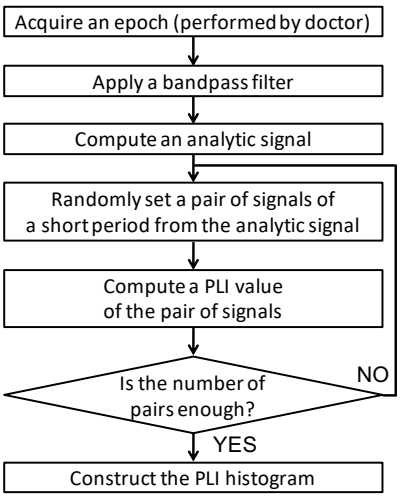

Fig. 1. Overview of generating our PLI histogram features.

provided by a doctor in a clinical setting. As described in [1], a doctor diagnoses encephalopathy by checking the functional connectivity of the brain. Specifically, the doctor observes a pair of EEG signals over a short period, acquired from distant brain regions. To automatically extract features for identifying epileptic seizure signals, using knowledge of the diagnosis, our method computes PLI values representing the functional connectivity of the brain from pairs of signals over short periods, which are randomly sampled from long-duration recordings of EEG signals. It then constructs histograms using the PLI values acquired from the pairs of signals, as shown in Fig. 1. The details of our feature extraction method are described below.

B. Procedure

To view EEG signals from an epilepsy patient, a doctor uses an EEG monitoring system that has C connection channels. The doctor acquires an epoch of multi-channel EEG signals with the time length T_1. First, our method simply applies a bandpass filter to an epoch to extract signals in specific frequency bands. This frequency band is specified as $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$, where ($\delta$: 0.5−4 Hz), ($\theta$: 4−8 Hz), ($\alpha$: 8−13 Hz), ($\beta$: 13−30 Hz), and ($\gamma$: above 30 Hz). Suppose that we acquire an EEG signal $x_{b,c}(t)$ at discrete time $t \in \{1, \cdots, T_1\}$ in the c-th channel $c \in \{1, \cdots, C\}$ of an epoch. To compute the PLI values, we generate an analytic signal $z_{b,c}(t)$ of the c-th channel of frequency band b, as

$$z_{b,c}(t) = x_{b,c}(t) + j\,\mathrm{H}(x_{b,c}(t))\ , \tag{2}$$

where j is the imaginary unit and $\mathrm{H}(x_{b,c}(t))$ represents the Hilbert transform of $x_{b,c}(t)$.

To compare signals over a short period in the analytic signal $z_{b,c}(t)$, we use the pair of signals acquired during a period of length $T_2 (< T_1)$. Our method performs random sampling until S pairs are acquired, by repeating the following procedure. We denote the start time $(m, n)_k$ of the k-th pair $k \in \{1, \cdots, S\}$ by

$$m \sim \mathrm{U}(0, T_1 - T_2), \tag{3}$$
$$n \sim \mathrm{U}(0, T_1 - T_2), \tag{4}$$

where $\mathrm{U}(0, T_1 - T_2)$ represents the discrete uniform distribution that randomly generates a duration between zero and $T_1 - T_2$.

We define the pair of signals, over a short period, between the c-th channel and d-th channel ($c \neq d$) using the start time $(m, n)_k$ as

$$z^m(t') = z_{b,c}(t_c), \tag{5}$$
$$z^n(t') = z_{b,d}(t_d), \tag{6}$$

where $t' \in \{1, \cdots, T_2\}$ is a discrete time, $t_c = m + t' - 1$, $t_d = n + t' - 1$, and $c, d \in \{1, \cdots, C\}$ are the channels of the EEG signals.

Having randomly sampled S pairs of the short-period signals $z^m(t')$ and $z^n(t')$ with start time $(m, n)_k$, we compute a PLI value for each. A PLI value $p^{m,n}_{b,c,d}$, relating $z^m(t')$ and $z^n(t')$, is defined as

$$p^{m,n}_{b,c,d} = \left| \frac{1}{T_2} \sum_{t'=1}^{T_2} \mathrm{sgn}\left(\Delta\phi(t')\right) \right|, \tag{7}$$

$$\Delta\phi(t') = \arg\left(\frac{z^m(t')\overline{z^n(t')}}{\|z^m(t')\|\|z^n(t')\|} \right), \tag{8}$$

where $\arg(z)$ represents the argument of the complex number z and $\overline{z^n(t')}$ represents the complex conjugate of $z^n(t')$.

To generate the feature vector for identification, we compute the histogram of the PLI values acquired by the random sampling. We define the feature vector as the PLI histogram $\boldsymbol{h}$, to temporally and spatially compare the short-period signals between the different channels. $\boldsymbol{h}$ is defined as

$$\boldsymbol{h} = \sum_{c=1}^{C} \sum_{d=c+1}^{C} \left[h^1_{b,c,d}, h^2_{b,c,d}, \ldots, h^N_{b,c,d}\right]^{\mathrm{T}}, \tag{9}$$

where N is the number of bins in the histogram and $h^l_{b,c,d}$ is the frequency of the l-th bin. We count the frequency $h^l_{b,c,d}$ using a set consisting of $p^{m,n}_{b,c,d}$, defined in Eq. (7), as

$$h^l_{b,c,d} = \mathrm{card}\{p^{m,n}_{b,c,d} | a_l \leq p^{m,n}_{b,c,d} < a_{l+1}\}, \tag{10}$$

where card is an operator representing the number of elements in the set consisting of $p^{m,n}_{b,c,d}$ under the condition $a_l \leq p^{m,n}_{b,c,d} < a_{l+1}$. It is worth noting that $a_l = -1 + 2(l-1)/N$. When $p^{m,n}_{b,c,d} = 1$, we count it in $h^N_{b,c,d}$. The total number of pairs of short-period signals is CS. We apply L1-normalization to ensure that $||\boldsymbol{h}||_1 = 1$.

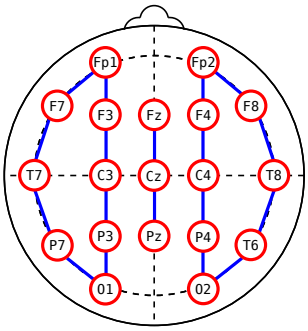

Fig. 2. Electrode alignment for recording EEG signals for bipolar EEG lead. The blue line is the connection channel between electrodes.

TABLE I
ACCURACY (%) OF EPILEPTIC SEIZURE IDENTIFICATION USING PLI HISTOGRAM FEATURES WHILE VARYING THE FREQUENCY BAND, LENGTH OF THE SHORT PERIOD, AND NUMBER OF BINS

Frequency band b				
δ	θ	α	β	γ
62.7±9.6	58.7±8.5	51.9±8.1	54.9±7.8	48.5±6.7

Length of short period T_2 (s)		
0.1	1	10
53.2±6.5	54.9±10.7	**58.1±10.4**

Number of bins N		
2	6	10
55.2±8.0	55.6±10.9	55.4±10.6

IV. EXPERIMENTS

A. Dataset

To evaluate the accuracy of distinguishing between epileptic seizure and non-seizure EEG signals using our PLI histogram features, we used the publicly available CHB-MIT Scalp EEG database [7]. This dataset contains EEG signals from 22 epilepsy patients (five males and 17 females, from 1.5 to 19 years old). The signals were annotated with the start times of epileptic seizures. We set the time length of each epoch $T_1 = 20$ s. The sampling rate was 256 Hz. We had to solve the problem that some EEG signals in the CHB-MIT Scalp EEG database were recorded using a different electrode alignment, so that they could not be used. To solve this problem, we selected only EEG signals recorded using the electrode alignment shown in Fig. 2. This alignment had $C = 18$ connection channels. In addition, we removed the patients whose signals contained fewer than seven epochs annotated as seizures. To remove the effects of artifacts, we did not use epochs with amplitudes exceeding 350 μV. Finally, we used EEG signals of 11 epilepsy patients (two male and nine female, from 1.5 to 16 years old). For each patient, both seizure and non-seizure EEG signals were recorded.

B. Experimental Results

We used the leave-one-patient-out cross-validation scheme to evaluate the accuracy of distinguishing between epileptic seizure and non-seizure EEG signals. Epochs of multi-channel EEG signals for one patient were used as test samples, and those for the remaining ten patients were used as training samples. Using the linear support vector machine (SVM), we compared the accuracy of the identification methods. We conducted our evaluations with frequency bands for $b = \delta$, θ, α, β, and γ signals, lengths of periods $T_2 = 0.1$, 1, and 10 s, and numbers of bins $N = 2$, 6, and 10. We randomly sampled $S = 1,000$ pairs of short-period signals. In the process of training the linear SVM classifier, we assigned epochs of seizure EEG signals to positive samples, and epochs of non-seizure EEG signals to negative samples.

Table I shows the identification accuracy of our PLI histogram features while varying the frequency band b, the length of the short period T_2, and the number of bins N. We calculated the average and standard deviation of the accuracy values obtained when one parameter was fixed and the others were changed. We can observe that the δ frequency band yielded higher accuracy than the θ, α, β, and γ frequency bands, and $T_2 = 10$ s yielded higher accuracy than $T_2 = 0.1$ s and 1 s. For the number of bins N, the accuracy was almost the same for $N = 2, 6$, and 10. We obtained the highest accuracy, 73.2% (sensitivity 58.2% and specificity 88.2%), using our features with the δ frequency band, $T_2 = 10$ s, and $N = 6$.

C. Visualization

We investigated which PLI histogram features contributed to enhancing the accuracy of identifying epileptic seizure EEG signals. We used the parameters of the feature achieving the highest accuracy, as described in Section IV-B. Figs. 3(a), (b), and (c) show the PLI histograms using δ signals for seizures and (d), (e), and (f) show those for non-seizures. The PLI values of seizure signals of short periods in the δ band were greater than zero more often than those of non-seizure signals. The pairs of signals taking large PLI values contained discriminative characteristics in δ waves. We believe that there is a high probability of phase synchronization in the δ band of the seizure EEG signals. This result may suggest an association with the appearance of the spike-and-slow-wave complex at about 3 Hz during seizures. Figs. 4(a), (b), and (c) show the PLI histograms using β signals for seizures, and (d), (e), and (f) show those for non-seizures. The PLI values of non-seizure signals of short periods in the β band frequently became one. The pairs of signals whose value was one contained discriminative characteristics in β waves. We believe that the phase synchronization appeared in non-seizure EEG signals of short periods in the β band. Because these results are not explained by existing medical knowledge, we need to perform further investigation.

D. Comparison with Cross-Correlation Features

We compared the accuracy of our PLI histogram features with prevalent cross-correlation features [8] using the five elements of the cross-correlation values, computed from the EEG signals between channels. The prevalent cross-correlation feature consisted of $5 \times 153 = 765$ dimensions because it

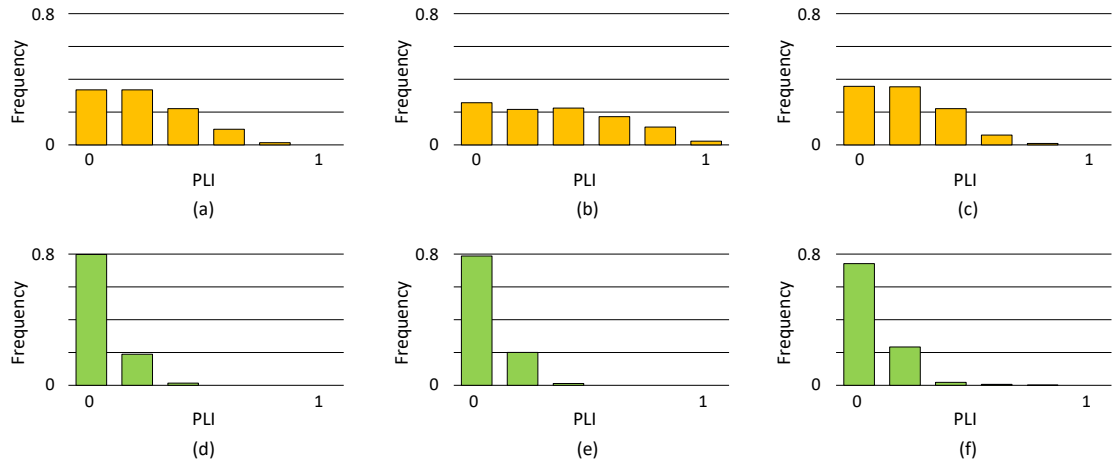

Fig. 3. Examples of PLI histograms of δ signals of patients experiencing epileptic seizures in (a), (b), and (c), and non-seizures in (d), (e), and (f).

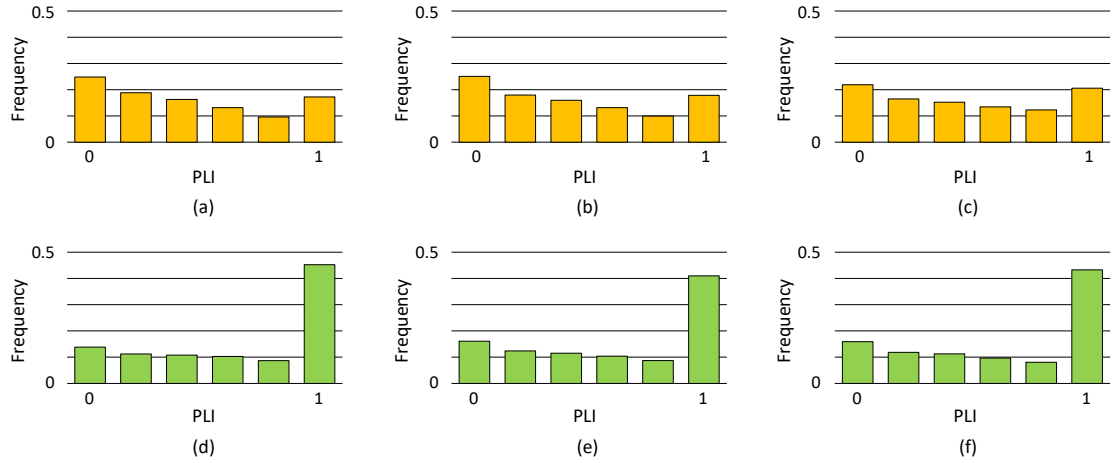

Fig. 4. Examples of PLI histograms of β signals of patients experiencing epileptic seizures in (a), (b), and (c), and non-seizures in (d), (e), and (f).

was computed from the five elements between ${}_{18}\mathrm{C}_2 = 153$ combinations of channels. Both features were trained using a linear SVM classifier. We used the same dataset as described in Section IV-A. Our PLI histogram features achieved an accuracy of 73.2%, whereas the prevalent cross-correlation features achieved 57.9%. Thus, our PLI histogram features improved the accuracy of identifying epileptic seizure EEG signals, compared with the prevalent cross-correlation features.

V. Conclusions

We proposed a method for identifying epileptic seizure EEG signals using PLI histogram features. By experiments on the CHB-MIT Scalp EEG database, we demonstrated that our PLI histogram features improved identification accuracy. In future work, we intend to evaluate the accuracy with a greater number of epilepsy patients and analyze the experimental results from a medical perspective.

References

[1] J. S. Ebersole, A. M. Husain, and D. R. Nordli, Eds., *Current Practice of Clinical Electroencephalography*, LWW, 2014.

[2] A. B. Waites, R. S. Briellmann, M. M. Saling, D. F. Abbott, and G. D. Jackson, "Functional connectivity networks are disrupted in left temporal lobe epilepsy," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 59, no. 2, pp. 335–343, 2006.

[3] W. Liao, Z. Zhang, Z. Pan, D. Mantini, J. Ding, X. Duan, C. Luo, G. Lu, and H. Chen, "Altered functional connectivity and small-world in mesial temporal lobe epilepsy," *PloS one*, vol. 5, no. 1, pp. e8525, 2010.

[4] M. Hardmeier, F. Hatz, H. Bousleiman, C. Schindler, C. J. Stam, and P. Fuhr, "Reproducibility of functional connectivity and graph measures based on the phase lag index (pli) and weighted phase lag index (wpli) derived from high resolution eeg," *PloS one*, vol. 9, no. 10, pp. e108648, 2014.

[5] S. Kasakawa, T. Yamanishi, T. Takahashi, K. Ueno, M. Kikuchi, and H. Nishimura, "Approaches of phase lag index to eeg signals in alzheimer's disease from complex network analysis," in *Innovation in Medicine and Healthcare 2015*, 2015, pp. 459–468.

[6] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources," *Human brain mapping*, vol. 28, no. 11, pp. 1178–1193, 2007.

[7] A. H. Shoeb, *Application of machine learning to epileptic seizure onset detection and treatment*, Ph.D. thesis, Massachusetts Institute of Technology, 2009.

[8] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of eeg signals," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1329–1336, 2009.