Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models

Sheng-Fu Liang, *Member, IEEE*, Chin-En Kuo, Yu-Han Hu, Yu-Hsiang Pan, and Yung-Hung Wang

***Abstract*—In this paper, we propose an automatic sleep-scoring method combining multiscale entropy (MSE) and autoregressive (AR) models for single-channel EEG and to assess the performance of the method comparatively with manual scoring based on full polysomnograms. This is the first time that MSE has ever been applied to sleep scoring. All-night polysomnograms from 20 healthy individuals were scored using the Rechtschaffen and Kales rules. The developed method analyzed the EEG signals of C3-A2 for sleep staging. The results of automatic and manual scorings were compared on an epoch-by-epoch basis. A total of 8480 30-s sleep EEG epochs were measured and used for performance evaluation. The epoch-by-epoch comparison was made by classifying the EEG epochs into five states (Wake/REM/S1/S2/SWS) by the proposed method and manual scoring. The overall sensitivity and kappa coefficient of MSE alone are 76.9% and 0.65, respectively. Moreover, the overall sensitivity and kappa coefficient of our proposed method of integrating MSE, AR models, and a smoothing process can reach the sensitivity level of 88.1% and 0.81, respectively. Our results show that MSE is a useful and representative feature for sleep staging. It has high accuracy and good home-care applicability because a single EEG channel is used for sleep staging.**

***Index Terms*—Automatic sleep scoring, autoregressive (AR) model, linear discriminant analysis (LDA), multiscale entropy (MSE), single-channel electroencephalogram (EEG).**

I. Introduction

HUMAN beings spend approximately one-third of their lives sleeping. Sleep diseases, such as insomnia and obstructive sleep apnea, seriously affect patients' quality of life. The prevalence of insomnia symptoms without restrictive criteria is approximately 33% in the general population [1]. Obstructive sleep apnea affects over 2% of adult women and 4% of adult men [2]. These sleep problems may cause daytime sleepiness, irritability, depressive or anxious mood, or even death.

Manuscript received May 19, 2011; revised October 11, 2011; accepted January 5, 2012. Date of publication March 6, 2012; date of current version May 11, 2012. This work was supported by the National Science Council of Taiwan under Grants NSC 98-2221-E-006-161-MY3 and NSC 100-2220-E-006-010. The Associate Editor coordinating the review process for this paper was Dr. Jiong Tang.

S.-F. Liang is with the Department of Computer Science and Information Engineering and the Institute of Medical Informatics, National Cheng Kung University (NCKU), Tainan 701, Taiwan (e-mail: sfliang@mail.ncku.edu.tw).

C.-E. Kuo and Y.-H. Hu are with the Department of Computer Science and Information Engineering, NCKU, Tainan 701, Taiwan (e-mail: chihen.kuo@gmail.com; hangohu@gmail.com).

Y.-H. Pan is with Ledder Technologies Holdings Ltd., Taipei 241, Taiwan (e-mail: D95310001@mail.ntou.edu.tw).

Y.-H. Wang is with the Society of Streams, Kaohsiung 443, Taiwan.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2012.2187242

For diagnosis of sleep problems, all-night polysomnographic (PSG) recordings including electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) are usually taken from the patients and scored by a well-trained expert according to the Rechtschaffen and Kales (R&K) rules [3] to classify each epoch (i.e., 30-s data) into one of the sleep stages, including wakefulness (Wake), non-rapid eye movement (NREM; stages 1–4), and REM. Recently, stages 3 and 4 were combined to become the new slow-wave sleep stage (SWS).

Because visual sleep scoring is a time-consuming and subjective process, automatic sleep-staging methods based on multichannel signals, including EEG, EMG, and EOG [4]–[13] or single EEG channel [10], [14], were developed. These methods contain two processes: feature extraction to analyze the recording epoch and classification to identify the sleep stage of the epoch. According to the R&K standard, some features have been proposed for sleep staging, including the alpha ratio [9], spindle ratio [12], and SWS ratio [13]. In addition, spectral power, power ratio, and spectral frequency [7] have been used in previous methods. In classification, many methods have been proposed, such as linear discriminant analysis (LDA) [15], artificial neural network [7], fuzzy system [10], and decision tree [16]. The overall agreements of these methods were in the range of 80%–85%.

According to the R&K rules, the EEG, EOG, and EMG signals are required for the expert to score the sleep stages. Therefore, multichannel PSG signals were utilized for most of the automatic sleep-staging methods [4]–[13]. However, the excessive number of wired connections for conventional polysomnography (PSG) is often a problem that leads to sleep disturbance. Automatic sleep-staging methods based on the single-channel EEG can reduce sleep disturbance caused by recording wires. The reported overall agreements of the single-channel-based methods are still less than 83% [10], [14]. The accuracy needs to be improved for real-world applications.

Recently, a new signal analysis method called multiscale entropy (MSE) has been proposed [17], [18] to estimate the complexity associated with the long-range temporal correlation of a time series. Instead of a single scale, it calculates the entropy of a time series over multiple temporal scales. MSE has been applied to analyze the complexity of various biomedical signals such as EEG [19]–[21], ECG [18], and heart rate [22], [23]. These studies show that MSE values of some biological signals are different between both patients and normal subjects and younger and older subjects.

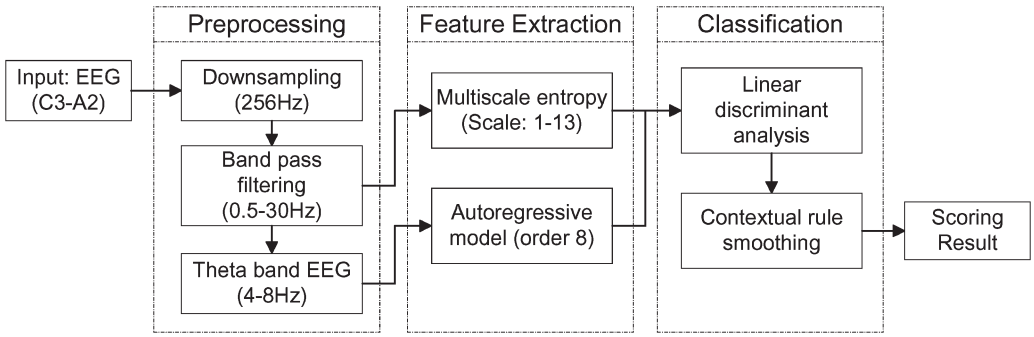

Fig. 1. Flowchart of single-channel sleep-staging method.

In this paper, MSE was first applied to analyze the EEG signals of different sleep stages to determine if it can be utilized as a new feature of sleep EEG analysis. An automatic sleep-scoring method for single-channel sleep EEG was then developed to combine MSE and autoregressive (AR) models [27] of EEG epochs from C3-A2 as features, and LDA [15] was utilized to classify each epoch into five sleep stages. The epoch-by-epoch comparison between the results of the proposed method based on signal-channel EEG and the results of the manual scoring based on PSG was made to demonstrate the performance of the developed system.

II. Materials and Method

A. Subjects and Recordings

All-night PSG sleep recordings were obtained from 20 healthy subjects (12 males and 8 females, aged 21.2 $\pm$ 1.1 years) for MSE analysis and to develop an automatic sleep-scoring method. The subjects were interviewed about their sleep quality and medical history. None of them reported any history of neurological or psychological disorders. The PSG recordings of each subject, including six EEG channels (F3-A2, F4-A1, C3-A2, C4-A1, P3-A2, and P4-A1, according to the international 10–20 standard system), two EOG channels (the above right and below left outer canthus), and a chin EMG channel, were acquired through the Siesta 802 PSG (Compumedics, Inc.). The sampling rate was 1 kHz with 16-b resolution. The filter settings of the cutoff frequencies were 0.5–30 Hz for EEG/EOG and 5–100 Hz for EMG. As suggested by the R&K rules [3], these nine-channel signals were used for manual scoring and only the data of the C3-A2 EEG channel were used for the developed single-channel sleep-staging system.

B. Manual Scoring

The 20 PSGs were scored by two sleep experts who worked independently from each other and according to the R&K rules [3]. Each 30-s epoch was classified into one of the five sleep stages, including Wake, REM, stage 1 (S1), stage 2 (S2), and slow-wave sleep (SWS), as well as movement artifact. In our experiments, only epochs belonging to the five sleep stages were used, and movement artifact epochs were rejected [7], [10]. Agreement was 92.3% between two sleep experts. In this paper, we compared the automatic classification and the manual scoring of expert 1. The results of comparison between expert 2 and automatic analysis did not differ statistically from those computed with expert 1.

C. Single-Channel Sleep-Staging System

Fig. 1 shows the flowchart of the proposed single-channel sleep-staging method that includes three parts: 1) preprocessing; 2) feature extraction; and 3) classification. The following figure presents each part in greater detail.

Part 1: Preprocessing: After downsampling the signals to 256 Hz for simplicity, an eighth-order Butterworth bandpass filter with 0.5–30-Hz passband is used to filter the downsampled signals for MSE analysis. In addition, an eighth-order Butterworth bandpass filter with 4–8-Hz passband was also utilized to extract the theta-band components for the AR model [25] to complement the MSE in accurately recognizing stage 1 [3].

Part 2: Feature Extraction: The continuous filtered signals were segmented into nonoverlapping 30-s windows (called epochs) for feature extraction. The feature extraction part contained two major analyses: 1) MSE and 2) AR model. The MSE is the principal analysis for the developed method, and the AR model is the complementary feature to raise the classification accuracy of stage 1.

Part 2.1. MSE: MSE measures the complexity of a time series by taking into account the entropy with respect to multiple temporal scales. MSE can be computed from the different types of entropy with multiple coarse-grained sequences, such as approximate entropy (ApEn) [26], [27] and sample entropy (SampEn) [28]. Although the ApEn method is widely used, it is inherently biased because self-matches are incorrectly counted to avoid the occurrence of natural logarithm of zero in the calculation. In addition, it has been suggested that this method may be heavily dependent on data length and that it lacks relative consistency. To overcome these limitations, SampEn was proposed by Richman and Moorman [28]. Therefore, SampEn was utilized as the kernel for entropy calculation in this paper.

Given an EEG time sequence with N samples, $x = \{x_1, x_2, \ldots, x_N\}$, the original times series is divided into nonoverlapping windows of length τ, defined as the scale factor.

The data points inside each window are then averaged. Each element of the coarse-gained time series $y_\tau(j)$ is calculated by the following:

$$y_\tau(j) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \qquad 1 \le j \le \frac{N}{\tau}. \tag{1}$$

After obtaining each element of the coarse-gained time series for each scale τ, we calculate the SampEn $SpEn(r, m, N/\tau)$ [28] for each of the coarse-grained time series. The parameter r is defined as the tolerance for accepting matches, and m is the dimension of the sequence vector. Some theoretical and clinical applications have shown that the parameters $m = 1$ or 2 and $r = 0.1 - 0.25$ of the standard deviation (SD) of the original time sequence provide good statistical validity for SampEn [29], [30]. The parameters $m = 2$ and $r = 0.15 * SD$ were used to calculate MSE values in this paper.

Part 2.2. AR Model: The AR model is a parametric model used to describe a stationary time series, and it is a popular tool for EEG analysis [10], [25], [31]. The model parameters can be used to determine the EEG states. AR models represent the current signal $x(t)$ as the weighted sum of its previous values $x(t-i)$ and the uncorrelated error $\varepsilon(t)$

$$x(t) = \sum_{i=1}^{p} a_i x(t-i) + \varepsilon(t) \tag{2}$$

where $a(i)$ represents the AR coefficients and p is the order of the AR model.

In this paper, the order of the AR model is eight, and the inputs of the AR model are the theta-band signals (4–8 Hz) extracted by an eighth-order Butterworth bandpass filter in the preprocessing. Combining the 8 AR coefficients and 13 MSE values, there is a total of 21 features in our method.

Part 3: Classification: A linear classifier, namely, LDA, was utilized to classify the extracted MSE values and AR coefficients into five sleep stages. In addition to reducing the computational cost, it can demonstrate the distinguishability of the proposed EEG features by using a linear classifier. Finally, the temporal contextual rules considering continuity, stage descending, and stage arousing [11], [33] were applied to smooth and fine-tune the classification results of LDA.

Part 3.1: LDA: LDA uses a hyperplane to determine the linear combination of features that best separates two or more classes of objects or events. Usually, the within-class, between-class, and mixture scatter matrices are used to formulate the criteria for searching the hyperplane so that the distance between the classes' means is minimized and the interclass variance is maximized [34], [35].

In order to demonstrate the generalization of the proposed method, the PSG data of ten subjects were used to train the LDA classifier, and the PSG data of the other ten subjects were used to verify the performance of our proposed method.

Part 3.2: Smoothing: Sleep staging has periodicity and continuity from light to deep [3]. After classifying the sleep stage by LDA, some misclassified epochs can be corrected according to the temporal contextual information and R&K rules in order to consider the temporal contextual information. These rules refer to the relation between epochs prior to and posterior to the current epoch. For example, three consecutive epochs of S2, REM, S2 should be replaced with the sequence S2, S2, S2. Similarly, consecutive epochs of REM, S1, REM should be replaced with the sequence REM, REM, REM. According to the rules presented in [11] and [33], a total of 11 rules were utilized to smooth the final results and increase the accuracy of our method.

TABLE I
TOTAL NUMBER OF 30-s EEG EPOCHS FROM 20 SUBJECTS USED FOR TRAINING (FROM TEN SUBJECTS) AND TESTING (FROM THE OTHER TEN SUBJECTS) THE PROPOSED METHOD

	Epochs	Wake	S1	S2	SWS	REM
Training	8496	3.08 %	4.40%	54.41%	17.78%	20.32%
Testing	8480	2.67 %	2.98 %	54.59 %	17.81 %	21.96 %

III. RESULTS

The single-channel EEG signals (C3-A2) from ten subjects were used to train the proposed model, and the EEG data of the other ten subjects were used to test the constructed model. Table I shows a total number of 30-s EEG epochs used for training and testing, and a percentage of the epochs corresponding to each sleep stage was also given. The numbers of epochs for training and testing are 8496 and 8480, respectively.

The experiments consist of three parts: 1) investigation of the relation between MSE values and sleep stages; 2) performance evaluation of the developed method; and 3) performance comparison of the developed method and the related methods.

A. Relation Between MSE Values and Sleep Stages

Fig. 2 shows the relation between MSE values and sleep stages. Fig. 2(a) shows the MSE values of different scales of one subject (No. 9), and Fig. 2(b) shows the results of averaging the 20 MSE values ($\tau = 1-20$) in each epoch. Fig. 2(c) shows the manual sleep scoring by the expert. The correlation coefficient between the averaged MSE values and the manual scoring of sleep stages reaches a maximal level of 0.7628. The high positive correlation between MSE values and sleep stages is the mapping of the EEG complexity in different sleep stages and the corresponding MSE values. MSE is a measure index of signal regularity. Theoretically, if the complexity of the signal is greater, the entropy value will be higher. Relatively, the entropy value is smaller if the complexity of the signal is lower. According to Fig. 2(b) and (c), it can be observed that the levels of the EEG complexity decrease from wake to deep sleep, so the MSE values reflect the changes in EEG complexity. This encouraging result motivates us to apply the MSE to the development of automatic sleep-staging methods.

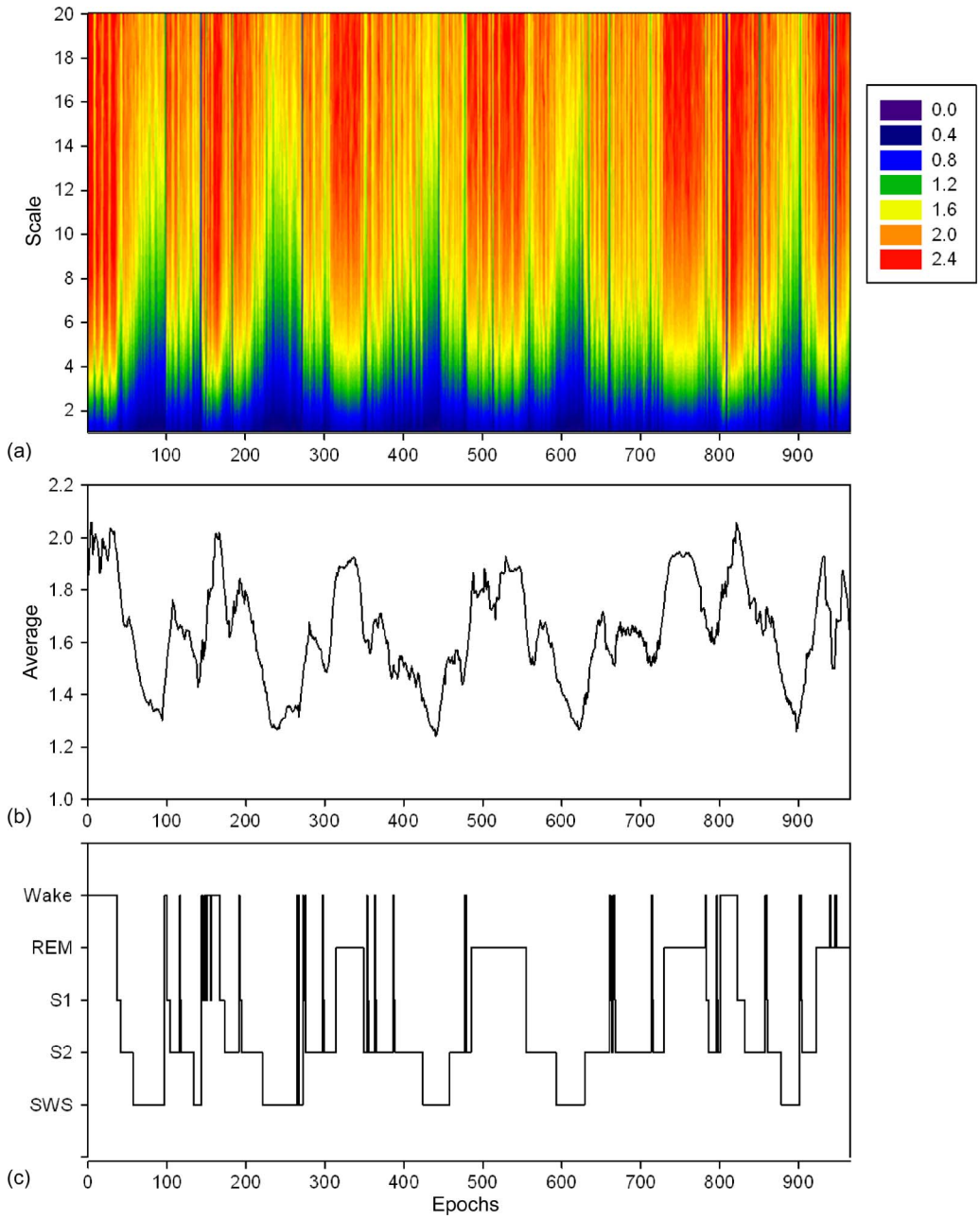

Fig. 2. Relation between MSE values and sleep stages. (a) Values of sleep EEG from subject No. 9. (b) Results of averaging the 20 MSE values in each epoch. (c) Manual sleep staging reviewed by the expert.

The mean MSE values $(\tau = 1{-}20)$ and SD from 20 subjects corresponding to different sleep stages are shown in Fig. 3. Some interesting characteristics can be observed: 1) In addition to Wake, the SampEn value increases when the scale factor increases for each stage; 2) for each scale factor, the entropy values and the levels of sleep depth have a negative correlation, and the entropy values monotonically decrease from wake to deep sleep (SWS); 3) the curves of S1 and REM overlap; and 4) when the scale factor is larger than 13, the SampEn values of each stage gradually flatten.

One-way repeated-measure ANOVA was applied to the data shown in Fig. 3 for statistical analysis. The results show that the SampEn values of any two stages are significantly different $(p < 0.05)$ for the scale factors of 1–8 and 20, except for differences between S1 and REM. For scale factors between 9 and 19, the SampEn values of any two stages are significantly different $(p < 0.05)$, including the differences between S1 and REM. However, the overlaps between Wake and REM and overlaps between S2 and SWS increase when the scale factor is large than 13. Based on the statistical analysis and results shown in Fig. 3, the MSE values corresponding to scales 1–13 were used as the principal features for our automatic sleep-staging system. Because the MSE values of S1 and REM are very close, the AR model was utilized as a complementary feature to enhance the performance of our system.

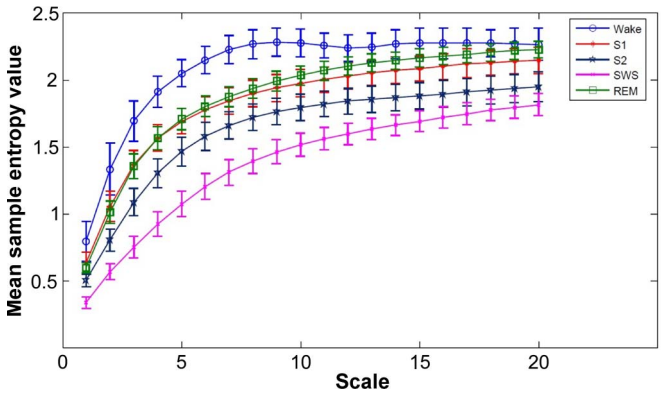

Fig. 3. MSE curves of EEG data (C3-A2) derived from 20 subjects with five different sleep stages: Wake, S1, S2, SWS1, and REM. SampEn ($m = 2, r = 0.15, N = 7680$) was evaluated at 20 scale factors. The symbols represent the mean values of SampEn for each stage, and the error bars represent the standard error, meaning the SD divided by the square root of the number of epochs.

TABLE II
CONFUSION MATRICES BETWEEN THE AUTOMATIC SCORING METHOD USING MSES AS FEATURES AND THE MANUAL SLEEP SCORING

		Computer						
		Wake	S1	S2	SWS	REM	Total	SE
Human	Wake	**200**	11	4	0	11	226	88.49%
	S1	67	**32**	25	3	126	253	12.46%
	S2	256	105	**3278**	310	680	4629	70.81%
	SWS	9	0	219	**1281**	1	1510	84.83%
	REM	42	15	67	7	**1731**	1862	92.96%
	Average SE						8480	**76.91%**

B. Performance Evaluation of the Developed Method

1) Performance of MSE on Sleep Staging: In order to determine the performance of MSE on sleep staging, only the MSE values (scale factors 1–13) of the single-channel EEG were used as features of the LDA classifier. Table II shows the confusion matrices of five-stage epoch classification by automatic staging versus manual scoring. The sensitivity [36] of computer scoring corresponding to each stage, and the average sensitivities are given. In addition, Cohen's kappa coefficient [37] was also calculated to assess the robustness of our system. Cohen's kappa coefficient (κ) is a statistical measure of inter-rater agreement among two or more raters. It is generally thought to be a more robust measure than simple percent agreement calculations because κ takes into account agreements that occur by chance. The interpretation of kappa coefficients by Landis and Koch is as follows: Values less than 0.00 indicate poor agreement; 0.00–0.20 indicate slight agreement; 0.21–0.40 indicate fair agreement; 0.41–0.60 indicate moderate agreement; 0.61–0.80 indicate substantial agreement; and those more than 0.80 indicate excellent agreement. The average sensitivity and kappa coefficient can reach up to 76.91% and 0.65, respectively. It shows that MSE is a good feature for sleep staging. Apart from S1 and S2, the sensitivities of the other stages are more than 84%.

Fig. 4 shows the sensitivity curves of each stage and the average result by using different numbers of MSE scales as features for sleep scoring. It was found that the sensitivity curve of the average results rose from approximately 50% to 77% when the numbers of scale factors used as features increased from 1 to 13. The average sensitivity is higher than 70% when the number of scale factors used as features is more than two. When the number of selected scale factors approaches 13, the average sensitivity approaches complete flatness. Although single-scale SampEn shows better performance on the sensitivities of SWS, MSE is more distinguishable and stable for overall performance. It is noted that the agreement of S1 is still less than 13% even when 13 scale factors were used. Therefore, the AR model was included as the complementary feature for improvement.

Agreement of AR Model Coefficients With Different Orders: In addition to the MSE values, the coefficients of the eighth-order AR model applied to the theta-band EEG signals (4–8 Hz) were extracted as the complementary features. Combining 8 AR coefficients and 13 MSE values, there are a total of 21 features in our method. Table III shows the confusion matrices of five-stage epoch classification by automatic staging 13 MSE values and 8 AR coefficients as features versus manual scoring. Compared with Table II, the average sensitivity increases from 76.91% to 85.38%. The sensitivities of S1, S2, and REM also increase by 9.27%, 14.19%, and 8.47%, respectively. In addition, the kappa coefficient is 0.77.

For better performance of sleep staging, we combine AR coefficients (orders 5–20) with 13 MSE values (the scale factor is 13) as features. In addition to increasing feature (AR coefficients 5–20), the remaining parts of the second experiment are the same as that of the first one. The sensitivity curves of AR coefficients with different orders overall and for each stage are shown in Fig. 5. It could be found that the maximum sensitivities of each stage are distributed in different AR coefficient orders, although the optimal order for overall sensitivity is eight, and the sensitivity curve of Wake is unstable with the increase of AR coefficients.

2) Results After Smoothing: In order to raise the sensitivity of each stage and overall, we utilize the smoothing rules mentioned earlier to smooth and fine-tune the results of the classifier. Finally, five-state hypnograms were obtained using our system. As an illustration example, Fig. 6(a)–(c) shows the hypnograms of a subject that was scored by the expert and generated by our proposed method without and with smoothing, respectively. Compared with the hypnogram of result without smoothing, we can see that the hypnogram with smoothing is more similar to the hypnogram scored by the expert.

The sensitivities of each stage between our proposed method with smoothing and manual scoring by the expert are shown in Table IV. The overall sensitivity and kappa coefficient are 88.11% and 0.81, respectively. Compared to Table III, the sensitivities of each stage and the overall sensitivity were all increased by smoothing. Clearly, smoothing can help the sensitivity of every stage and overall sensitivity. Moreover, the performance of our system is very good in each stage except S1. The sensitivities of S1 and other stages are more than 28% and 86%, respectively.

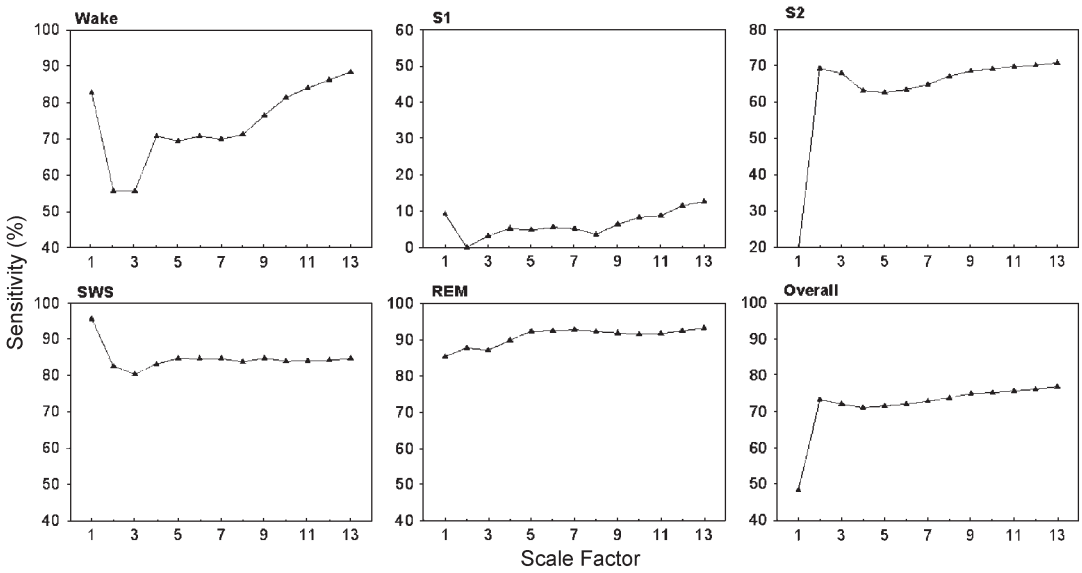

Fig. 4. Sensitivity curves of each stage and the average result by using different numbers of MSE scales as features for sleep scoring.

TABLE III
CONFUSION MATRICES BETWEEN THE AUTOMATIC SCORING METHOD INTEGRATING MSE AND AR AS FEATURES AND THE MANUAL SLEEP SCORING

		Computer						
		Wake	S1	S2	SWS	REM	Total	SE
Human	Wake	**191**	22	4	0	9	226	84.51%
	S1	27	**55**	42	3	126	253	21.73%
	S2	8	117	**3935**	277	292	4629	85.00%
	SWS	1	4	205	**1293**	7	1510	85.62%
	REM	25	20	44	6	**1767**	1862	94.89%
	Average SE						8480	**85.38%**

C. Comparisons

In order to demonstrate the usefulness of MSE, we also compared the classification performance of our method using AR coefficients alone versus using MSE and MSE + AR coefficients to highlight the contribution of MSE features. The AR models were applied to the theta-band EEG (AR1) and the raw EEG signals (AR2) for comparison. To be fair, other conditions such as the use of the smoothing technique are identical for all cases.

The results in Table V show that: 1) the average sensitivity of using MSE coefficients alone is better than AR1 and AR2 by 4.51% and 18.26%, respectively; 2) after smoothing, the average sensitivity by using MSE coefficients alone is better than AR1 and AR2 by 5.85% and 19.28%, respectively; 3) the sensitivities of using MSE coefficients are also more balance compared with the AR coefficients. In addition, comparing AR1, MSE, AR1 + MSE, and AR1 + MSE + smoothing, we can conclude that: 1) MSE enjoys significant leap on the SWS sensitivity; 2) MSE affects Wake detection performance (less than 5%) while being combined with AR1; and 3) smoothing provides only marginal improvement. These results demonstrate that MSE is a stable and representative feature for sleep staging.

The proposed sleep-scoring method was also tested on the Physionet database that provides sleep recordings and corresponding hypnograms in European data format [38]. It contains the recordings (16 bits, 100 Hz) from eight healthy subjects (21–35 years old) with no medication. Only the EEG signals of the Pz-Oz channel were used in our experiment. It is noted that the sampling rate and the recording sites of EEG (100 Hz, Pz-Oz) are different from the recordings for the development of our method (256 Hz, C3-A2), and we went through a training phase again to fine-tune the system parameters. After feature extraction (MSE scales 1–13 and AR orders 1–8), the data from four subjects were used to train the LDA classifier, and the data from the other four subjects were used to verify the performance of our proposed method. The smoothing process is also applied. The final results are shown in Table VI. The sensitivities of overall agreement, Wake, S1, S2, SWS, and REM are 83.6%, 91.99%, 18.75%, 70.19%, 99.11%, and 85.42%, respectively. The method proposed in [10] has also been tested on this data set, and only single-channel EEG signals were used. It was reported that the average sensitivity and the sensitivities of Wake and REM are 71.2%, 85.2%, and 63%, respectively. The sensitivities of S1, S2, and SWS were not reported [10]. The sensitivities of our method are superior to the results of the method in [10] for all reported stages.

IV. DISCUSSION

In this paper, a novel approach based on MSE analysis of a single EEG channel (C3-A2) for automated sleep scoring was developed. Until now, this is the first time that MSE was applied to automatic sleep scoring, although MSE has been applied to compare the complexity of EEG signals from different subject groups [19]–[21], [39]. The proposed method integrated MSE, AR models, LDA, and a smoothing process, and the overall sensitivity and kappa coefficient of the method applied to the EEG from 20 subjects reached 88.1% and 0.81, respectively.

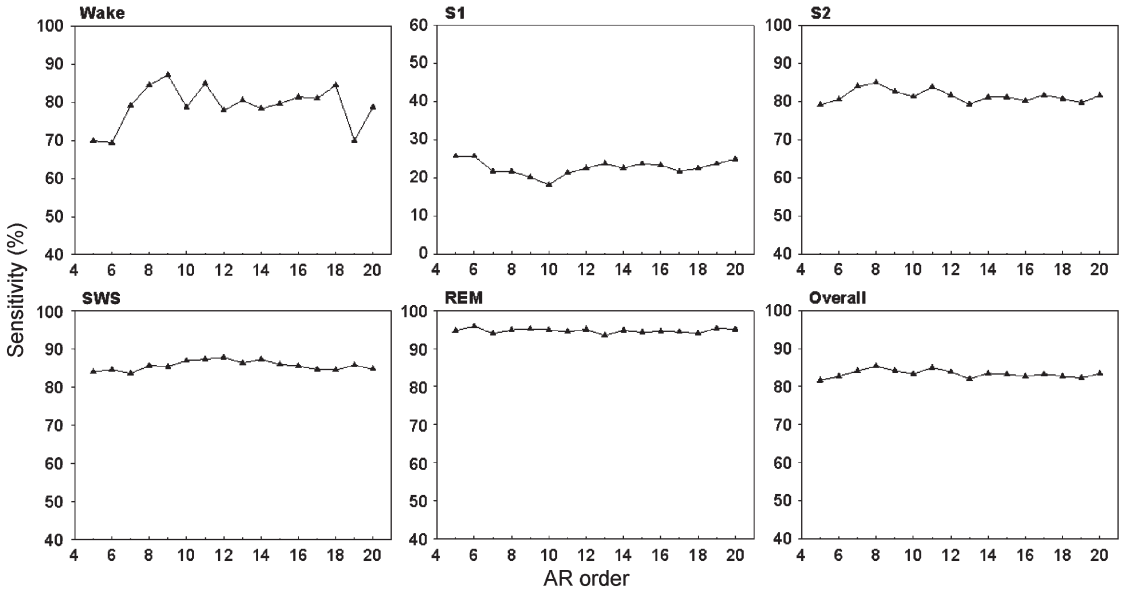

Fig. 5. Sensitivity curves of AR coefficients (orders 5–20) combined with MSE values (1–13) for each stage and overall.

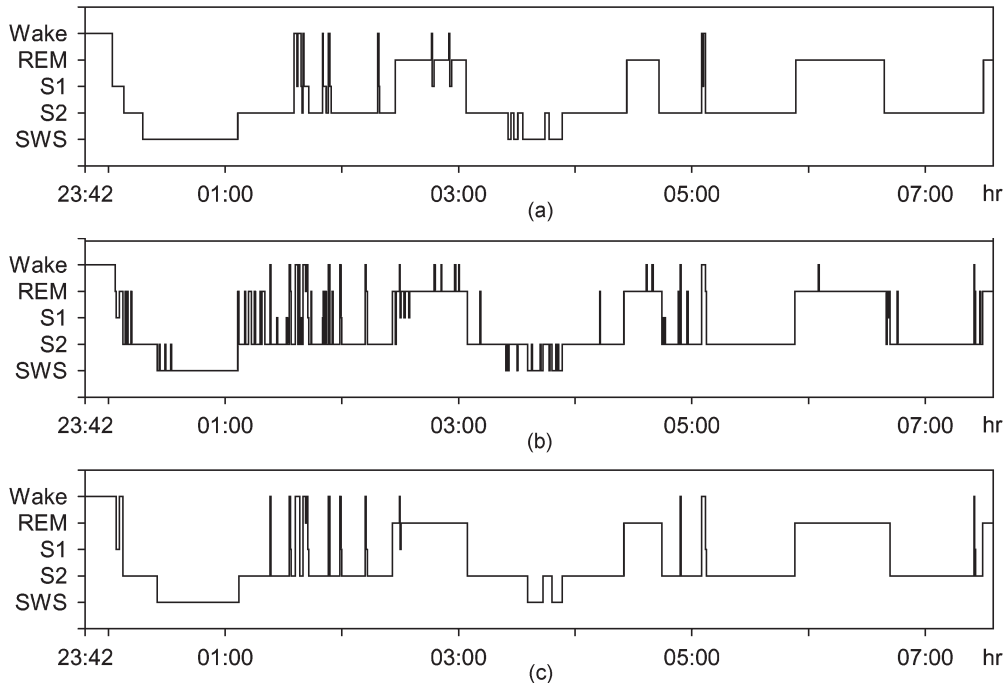

Fig. 6. Sleep hypnogram of the No. 9 subject in the testing set. (a) Scored by the expert. (b) Our proposed method without smoothing; sensitivity is 86.22%. (c) Our proposed method with smoothing; sensitivity is 89.13%.

Analyzing the relationship between the MSEs and sleep stages shows that the entropy values and the levels of sleep depth have a negative correlation for various scale factors. The MSE values monotonically decrease from wake to deep sleep in the same scale factor. The experimental results show that MSE is a powerful feature for sleep staging. Because the MSE values of S1 and REM are very close, the complementary features were required for the development of an automatic sleep-scoring system.

Compared to conventional sleep-scoring methods that require multiple physiological signals, including EEG, EOG, and EMG for feature extraction [8], [9], [11], the single-channel sleep-scoring approach has the advantage of reducing interference in sleep quality due to the use of fewer electrodes and the simplification of the preparation procedure. Recently, some single-channel automated sleep-staging methods have been developed [10], [14]. The reported sensitivities were in the range of 74%–82.9%, and our method is superior to these methods. However, because S1 and REM exhibit similar EEG patterns and S1 is a transition phase between Wake and the different sleep stages [40], most sleep-staging methods, including the proposed approach, have relatively low sensitivities in S1. Multichannel approaches also suffer from this problem [11].

TABLE IV
CONFUSION MATRICES BETWEEN THE AUTOMATIC SLEEP SCORING AFTER SMOOTHING AND THE MANUAL SLEEP SCORING (1–13 MSE VALUES AND 8 AR COEFFICIENTS)

		Computer						
		Wake	S1	S2	SWS	REM	Total	SE
Human	Wake	**195**	24	4	0	3	226	86.28%
	S1	61	**72**	48	3	69	253	28.45%
	S2	12	103	**4078**	216	220	4629	88.09%
	SWS	1	4	196	**1309**	0	1510	86.68%
	REM	8	8	22	6	**1818**	1862	97.63%
	Average SE						8480	**88.11%**

TABLE V
CLASSIFICATION RESULTS OF USING AR COEFFICIENTS ALONE (order = 1−8), MSE COEFFICIENTS ALONE (scale = 1−13), AND THE COMBINATION OF MSE + AR COEFFICIENTS

	Wake (%)	S1 (%)	S2 (%)	SWS (%)	REM (%)	Overall (%)
AR1	92.92	12.64	86.95	6.02	95.7	72.4
AR2	72.12	23.32	63.05	20.59	81.74	58.65
MSE	88.49	12.64	70.81	84.83	92.96	76.91
MSE+AR1	84.51	21.73	85	85.62	94.89	85.38
AR1+Smoothing	95.57	18.97	89.30	5.49	97.26	74.19
AR2+Smoothing	75.66	24.50	65.67	21.85	83.24	60.76
MSE+Smoothing	92.47	17.78	74.33	85.96	96.40	80.04
MSE+AR1+smoothing	86.28	28.45	88.09	86.68	97.630	88.11

*AR1: coefficients extracted from filtered EEG (theta band); AR2: coefficient extracted from raw EEG.

TABLE VI
CONFUSION MATRICES BETWEEN THE AUTOMATIC SLEEP SCORING AFTER SMOOTHING AND THE MANUAL SLEEP SCORING ON THE PHYSIONET PUBLIC DATABASE (1–13 MSE VALUES AND 8 AR COEFFICIENTS)

		Computer						
		Wake	S1	S2	SWS	REM	Total	SE
Human	Wake	**1849**	87	59	4	11	2010	91.99%
	S1	69	**24**	12	3	20	128	18.75%
	S2	15	45	**669**	165	59	953	70.19%
	SWS	0	1	1	**224**	0	226	99.11%
	REM	7	29	16	5	**334**	391	85.42%
	Average SE						3708	**83.60%**

Because visual sleep scoring is a time-consuming and subjective process, the proposed method can assist the clinical staff to reduce the time required for sleep scoring in the future. Because coming in to the sleep laboratory for a PSG examination is inconvenient, a portable sleep quality monitoring system integrating an embedded system for EEG acquisition [41], [42] and the automatic sleep-scoring method with reduced channel requirement is a practical approach for home care. This approach can also reduce the interference in sleep quality due to the unfamiliar environment at sleep centers and provide more recording trials at a lower cost to average out the abnormalities that may be seen in a one-night recording. We plan to combine the proposed algorithm into a portable EEG recording device to provide more portability and wearability for sleep quality evaluation at home.

REFERENCES

[1] M. M. Ohayon, "Epidemiology of insomnia: What we know and what we still need to learn," *Sleep Med. Rev.*, vol. 6, no. 2, pp. 97–111, Apr. 2002.

[2] M. W. Mahowald and C. H. Schenck, "Insights from studying human sleep disorders," *Nature*, vol. 437, no. 7063, pp. 1279–1285, Oct. 2005.

[3] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington, DC: Nat. Health Inst., 1968.

[4] J. R. Smith, M. Negin, and A. H. Nevis, "Automatic analysis of sleep electroencephalograms by hybrid computation," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-5, no. 4, pp. 278–284, Oct. 1969.

[5] J. D. Smith and I. Karacan, "EEG sleep stage scoring by an automatic hybrid system," *Electroencephalogr. Clin. Neurophysiol.*, vol. 31, no. 3, pp. 231–237, Sep. 1971.

[6] H. Kuwahara, H. Higashi, Y. Mizuki, S. Matsunari, M. Tanaka, and K. Inanaga, "Automatic real-time analysis of human sleep stages by an interval histogram method," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 3, pp. 220–229, Sep. 1988.

[7] N. Schaltenbrand, R. Lengelle, M. Toussaint, R. Luthringer, G. Carelli, A. Jacqmin, E. Lainey, A. Muzet, and J. P. Macher, "Sleep stage scoring using the neural network model: Comparison between visual and automatic analysis in normal subjects and patients," *Sleep*, vol. 19, no. 1, pp. 26–35, Jan. 1996.

[8] H. J. Park, K. S. Park, and D. U. Jeong, "Hybrid neural-network and rule-based expert system for automatic sleep stage scoring," in *Proc. 22nd Annu. EMBS Int. Conf.*, 2000, pp. 1316–1319.

[9] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1412–1423, Dec. 2001.

[10] C. Berthomier, X. Drouot, M. Herman-Stoïca, P. Berthomier, J. Prado, D. Bokar-Thire, O. Benoit, J. Mattout, and M. P. d'Ortho, "Automatic analysis of single-channel sleep EEG: Validation in healthy individuals," *Sleep*, vol. 30, no. 11, pp. 1587–1595, Nov. 2007.

[11] J. Virkkala, J. Hasan, A. Värri, S. Himanen, and K. Müller, "Automatic sleep stage classification using two-channel electro-oculography," *J. Neurosci. Methods*, vol. 166, no. 1, pp. 109–115, Oct. 2007.

[12] F. Duman, A. Erdamar, O. Erogul, Z. Telatar, and S. Yetkin, "Efficient sleep spindle detection algorithm with decision tree," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 9980–9985, Aug. 2009.

[13] C. Berthomier, J. Prado, and O. Benoit, "Automatic sleep EEG analysis using filter banks," *Biomed. Sci. Instrum*, vol. 35, pp. 241–246, 1999.

[14] A. Flexer, G. Gruber, and G. Dorffner, "A reliable probabilistic sleep stager based on a single EEG signal," *Artif. Intell. Med.*, vol. 33, no. 3, pp. 199–207, Mar. 2005.

[15] K. Šušmáková and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, Nov. 2008.

[16] P. Anderer, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klosch, B. Saletu, J. Zeitlhofer, M. J. Barbanoj, H. Danker-Hopfe, S. L. Himanen, B. Kemp, T. Penzel, M. Grozinger, D. Kunz, P. Rappelsberger, A. Schlogl, and G. Dorffner, "An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: Validation study of the Somnolyzer 24 × 7 utilizing the Siesta database," *Neuropsychobiology*, vol. 51, no. 3, pp. 115–133, 2005.

[17] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, p. 068102, Aug. 2002.

[18] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E.*, vol. 71, no. 2, p. 021906, Feb. 2005.

[19] T. Takahashi, R. Y. Cho, T. Murata, T. Mizuno, M. Kikuchi, K. Mizukami, H. Kosaka, K. Takahashi, and Y. Wada, "Age-related variation in EEG complexity to photic stimulation: A multiscale entropy analysis," *Clin. Neurophys.*, vol. 120, no. 3, pp. 476–483, Mar. 2009.

[20] J. H. Park, S. Kim, C. H. Kim, A. Cichocki, and K. Kim, "Multiscale entropy analysis of EEG from patients under different pathological conditions," *Fractals-Complex Geom Patterns Scaling Nat. Soc.*, vol. 15, no. 4, pp. 399–404, Dec. 2007.
[21] X. Kang, X. Jia, R. G. Geocadin, N. V. Thakor, and A. Maybhate, "Multiscale entropy analysis of EEG for assessment of post-cardiac arrest neurological recovery under hypothermia in rats," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1023–1031, Apr. 2009.
[22] M. Costa, C. K. Peng, A. L. Goldberger, and J. M. Hausdorff, "Multiscale entropy analysis of human gait dynamics," *Phys. A.*, vol. 330, no. 1/2, pp. 53–60, Dec. 2003.
[23] P. R. Norris, S. M. Anderson, J. M. Jenkins, A. E. Williams, and J. A. Morris, Jr., "Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 trauma patients," *Shock*, vol. 30, no. 1, pp. 17–22, Jul. 2008.
[24] S. Ishiwa, T. Ogawa, and H. Sonoda, "Developmental characteristics of topographic EEG in the newborn using an autoregressive model," *Brain Topography*, vol. 4, no. 1, pp. 23–30, 1991.
[25] N. V. Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 453–495, Aug. 2004.
[26] S. M. Pincus, "Approximate entropy: A complexity measure for biological time series data," in *Proc. 1991 IEEE 17th Annu. Northeast, Bioengineering Conf.*, 1991, pp. 35–36.
[27] S. M. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos Interdisciplinary J. Nonl. Sci.*, vol. 5, no. 1, pp. 110–117, Mar. 1995.
[28] J. S. Richman and J. R. Moorman, "Physiologic time series analysis using approximate entropy and sample entropy," *Am J. Physiol Heart Circ.*, vol. 278, no. 6, pp. 2039–2049, Jun. 2000.
[29] N. Burioka, M. Miyata, G. Cornelissen, F. Halberg, T. Takeshima, D. T. Kaplan, H. Suyama, M. Endo, Y. Maegaki, T. Nomura, Y. Tomita, K. Nakashima, and E. Shimizu, "Approximate entropy in the electroencephalogram during wake and sleep," *Clin. EEG Neurosci.*, vol. 36, no. 1, pp. 21–24, Jan. 2005.
[30] D. R. Chialvo, "Physiology: Unhealthy surprises," *Nature*, vol. 419, no. 6904, p. 263, Sep. 2002.
[31] E. Olbrich, P. Achermann, and P. F. Meier, "Dynamics of human sleep EEG," *Neurocomputing*, vol. 52–54, pp. 857–862, Jun. 2003.
[32] J. Pardey, S. Roberts, and L. Tarassenko, "A review of parametric modelling techniques for EEG analysis," *Med. Eng. Phys.*, vol. 18, no. 1, pp. 2–11, Jan. 1996.
[33] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events*. Westchester, IL: Amer. Acad. Sleep Med., 2007.
[34] B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.
[35] C. T. Lin, K. L. Lin, L. W. Ko, S. F. Liang, B. C. Kuo, and I. F. Chung, "Nonparametric single-trial EEG feature extraction and classification of driver's cognitive responses," *EURASIP J. Appl. Signal Process.*, vol. 2008, pp. 849 040-1–849 040-10, 2008.
[36] D. G. Altman, *Practical Statistics for Medical Research*. London, U.K.: Chapman & Hall, 1991, pp. 409–419.
[37] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.
[38] International Database PhysioNet Sleep Recordings. [Online]. Available: http://www.physionet.org/physiobank/database/
[39] J. Escudero, D. Abasolo, R. Hornero, P. Espino, and M. Lopez, "Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy," *Physiol. Meas.*, vol. 27, no. 11, pp. 1091–1106, Nov. 2006.
[40] S. L. Himanen and J. Hasan, "Limitations of Rechtschaffen and Kales," *Sleep Med. Rev.*, vol. 4, no. 2, pp. 149–167, Apr. 2004.
[41] C. T. Lin, L. W. Ko, J. C. Chiou, J. R. Duann, R. S. Huang, S. F. Liang, T. W. Chiu, and T. P. Jung, "Noninvasive neural prostheses using mobile and wireless electroencephalography," *Proc. IEEE*, vol. 96, no. 7, pp. 1167–1183, Jul. 2008.
[42] C. P. Young, S. F. Liang, D. W. Chang, Y. C. Liao, F. Z. Shaw, and C. H. Hsieh, "A portable wireless on-line closed-loop seizure controller in freely moving rats," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 513–521, Feb. 2011.

Sheng-Fu Liang (M'09) was born in Tainan, Taiwan, in 1971. He received the B.S. and M.S. degrees in control engineering and the Ph.D. degree in electrical and control engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1994, 1996, and 2000, respectively.

From 2001 to 2005, he was a Research Assistant Professor with the Department of Electrical and Control Engineering, NCTU. He joined the Department of Biological Science and Technology, NCTU, in 2005 and also the Department of Computer Science and Information Engineering (CSIE) and the Institute of Medical Informatics (IMI), National Cheng Kung University (NCKU), Tainan, in 2006. He is currently an Associate Professor with CSIE and IMI, NCKU. He is also a Collaborative Researcher with the Biomimetic Systems Research Center, NCTU. His current research interests are neural engineering, biomedical engineering, biomedical signal/image processing, machine learning, and medical informatics.

Chih-En Kuo was born in Tainan, Taiwan, in 1983. He received the B.S. degree in mathematics from National Cheng Kung University (NCKU), Tainan, in 2005 and the M.S. degree in information management from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 2009. He is currently working toward the Ph.D. degree in the Department of Computer Science and Information Engineering, NCKU.

His research interests are biomedical signal processing, algorithm analysis, human sleep EEG analysis, and automatic sleep staging.

Yu-Han Hu was born in Kaohsiung, Taiwan, in 1985. He received the B.S. degree in computer science and information engineering from National Cheng Kung University (NCKU), Tainan, Taiwan, in 2007. He is currently working toward the M.S. degree in the Department of Computer Science and Information Engineering, NCKU.

His research interests are biomedical signal processing, brain–computer interface, human sleep EEG analysis, and automatic sleep staging.

Yu-Hsiang Pan received the Ph.D. degree in physical science from National Taiwan Ocean University, Keelung, Taiwan, in 2010.

He has been the President of Ledder Technologies Holdings, Ltd., Taipei, Taiwan since 2010. His research interests include LED general lighting applications, signal processing, bioacoustics, and multiscale entropy algorithm.

Yung-Hung Wang was born in Taipei, Taiwan, in 1965. He received the B.S. degree in naval architecture from National Taiwan University, Taipei, the M.S. degree in mechanical engineering from the University of California, Santa Barbara, and the Ph.D. degree in aeronautics and astronautics from Stanford University, Stanford, CA, in 1998.

He is currently with the Society of Streams, Taiwan. His research interests include computational physics, discrete mathematics, electronic design automation, and signal processing.