Sleep Stage Recognition using Respiration Signal

Jialei Yang, James M. Keller, *Life Fellow, IEEE*, Mihail Popescu, *Senior Member, IEEE* and Marjorie Skubic, *Senior Member, IEEE*

***Abstract*—This paper presents a sleep stage recognition system for Awake, rapid eye movement (REM) and non-REM (NREM) sleep detection. Two respiratory variability (RV) features are extracted from oro-nasal airflow signals provided in the sleep-EDF (Expanded) database. A two layer system with threshold comparison classifier is implemented. This system achieved state-of-the-art performance with simple features and classifiers. The average accuracy of 74.00% ±5.30% and Cohen's kappa coefficient of 0.49 ±0.08 were achieved with 21 recordings. In the end, the measure of sleep efficiency was calculated and the average absolute error was 3.61% ±3.66%.**

I. Introduction

Sleep is understood as a reversible state of unconsciousness, characterized by a decrease of activity and alertness [1]. It is an essential activity for humans to maintain health. The lack of sleep or low quality of sleep will affect normal activities and cause physical and mental issues. The ability of monitoring sleep quality continually can help find sleep disorders instantly. Furthermore, some research showed a relation between sleep and other diseases (Parkinson's disease [2], Alzheimer's disease [3]). Therefore, the study of sleep is highly important.

The criteria of sleep stages were first standardized in 1968 by Allan Rechtschaffen and Anthony Kales (R&K scoring manual) [4]. In 2004, the AASM commissioned a revision of sleep scoring rules, covering not only sleep stages but also the scoring of arousals, respiratory events, sleep related movement disorders and cardiac abnormalities [5]. The revised scoring manual was published in 2007 [6].

Both manuals divided sleep into two main stages: rapid eye movement (REM) and non-REM (NREM). Current gold standard sleep study for the evaluation of sleep stages is polysomnography (PSG). The system is usually performed at a sleep lab. Subjects need to connect to various physiological signals such as: electroencephalogram (EEG), electromyogram (EMG), electrooculogram (EOG), etc. A trained expert gives a sleep stage every 30 seconds based on specific patterns in the obtained signals. So the subjects require not only the connection of various sensors and electrodes but also must spend the night in a bed that is different from their own [7]. These settings are inconvenient and also may affect a subject's sleep patterns. Therefore, a home-use and more efficient system is required for long-term sleep monitoring.

J. Y., J. K., and M. S. are with the ECE Department., University of Missouri (e-mail:jyzd7@mail.missouri.edu, kellerj@missouri.edu, skubicm@missouri.edu).

M. P. is with the HMI Department., University of Missouri (e-mail:popescum@missouri.edu).

The Hydraulic Bed Sensor (MUHBS) proposed in [8] is a non-invasive sensor that can capture the ballistocardiograpy (BCG) signal. From the BCG signal, heart rate variability (HRV) and respiratory variability (RV) features can be computed. These two types of features have been applied to sleep analysis in many studies [9], [10],[11] and showed good results. Thus, the MUHBS is a good fit for sleep monitoring. However, because we can't get access to a sleep lab at this time, in order to further analyze the potential of recognizing sleep stages using RV parameters, respiration signals in the Sleep-EDF Database (Expanded) [12] were used to examine features and classifiers. It is a standard on-line dataset that was de-identified.

This paper presents a three stage (Awake, REM and NREM) classification algorithm using the respiration signal and derived RV features. The purpose of this research is to develop an algorithm that can be applied to the bed sensor data in the future.

II. METHODS

A. *Dataset Description*

One of the studies in the Sleep-EDF Database (Expanded) was used. There are twenty healthy subjects (ten males and ten females; age 25-34 years). Each subject was recorded on two subsequent day-night periods. Subjects wore a modified Walkman-like cassette tape recorder described in [13] for about 20 hours in their homes. Several signals were recorded including a respiration signal obtained from an oral-nasal respiration air flow [13]. The oral-nasal airflow signal was sampled at 1 Hz.

Sleep stages in each 30 second epoch were given by annotation files according to the R&K standard. These stages were further combined to Awake (Awake and motion time), REM and NREM (N1, N2, N3 and N4). Because each 20 hour recording contains both daily living activities and sleep, data during the sleep time need to be extracted. Epochs from 20 minutes before the first non-awake stage to the last non-awake stage were considered as sleep time. Respiratory signals and detected sleep stages in this range were kept and data outside this range were excluded.

B. *Feature Extraction*

The respiration signals were segmented to 30 second epochs in order to match the given sleep stage annotations. Two features were extracted from each respiration signal segment: the respiratory rate (RR) and the max absolute differences of breath intervals (MADI). Compared with other RV features, these two showed significant differences among different sleep stages.

First, peaks of breath cycles were detected by finding local maxima. A 30 second respiration signal and detected peaks are shown in Fig. 1. For each 30 second epoch, we then defined the positions of the peaks as x(n), where n is a series number from 1 to the number of peaks (N). The breath intervals are:

$$I(n) = x(n+1) - x(n) \quad n = 1, 2, ...N\text{-}1 \tag{1}$$

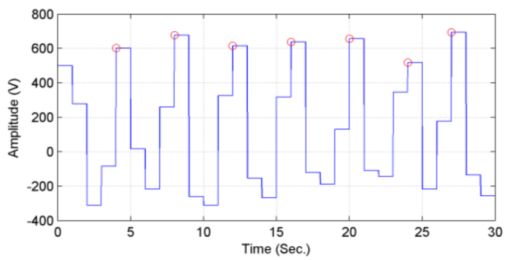

Figure 1. A 30 sec. respiration signal with detected peaks labeled with red circles

The equation for RR is:

$$\frac{60}{N}\sum_{n=1}^{N}\frac{1}{I(n)} \tag{2}$$

The MADI is defined as:

$$\max_{1<=n<=N-2}\left|I(n+1) - I(n)\right| \tag{3}$$

C. Data Selection

After analyzing all breath intervals of each recording, we found that some recordings are really noisy. Fig. 2 displays the histogram of breath intervals of one such recording. The breath intervals of this recording gather around 2 second which results in 30 breaths/minutes.

According to [14], a healthy adult breathes 12 - 15 times per minute at rest. So these recordings were considered as noisy data. In order to keep the whole dataset reliable, a data selection process was implemented to detect and remove these noisy recordings from the database. The selection rule is: for each recording, if the proportion of breath intervals larger than 6 seconds or smaller than 3 seconds is bigger than 10%, then this is a noisy recording.

This range selected the recordings where the majority of the respiratory rates are in the range of 10 to 20 times per minute, which allows some variability. After selection, 21 recordings were kept in the dataset. The structure of these recordings are listed in Table I.

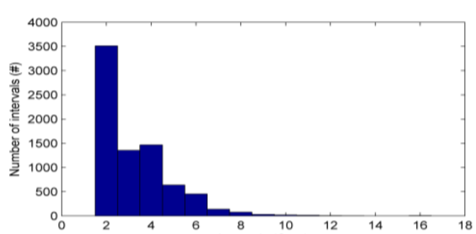

Figure 2. Histogram of breath intervals of one recording. Most of the intervals are 2 seconds

TABLE I. STRUCTURE OF THE DATABASE

Total number of epochs	20071
Awake (%)	9.90±3.90
REM (%)	20.48±3.91
NREM (%)	69.62±6.18

Mean±SD

III. SLEEP STAGE RECOGNITION SYSTEM

The system consists of two layers. The first layer was the Awake&REM detection. The second layer was to separate REM and Awake based on the outputs from layer one.

A. Layer 1 : Awake&REM detection

The RR feature was employed in the first layer. The RR series of all epochs was first detrended to remove the nonlinear trend. The detrended RR was calculated by subtracting the smoothed feature from the original one. The smoothed feature was obtained by applying RLOWESS [15] with a 30 point sliding window.

Thresholds were then set as the 3rd quartile value plus the standard deviation and the 1st quartile value minus the standard deviation. All epochs that had the detrended RR value out of range of these two thresholds were classified as Awake&REM class (detected target class). Fig. 3 displays the original RR, the detrended RR with thresholds and the ground truth with detected epochs.

In addition, two rules based on common sense and theory were applied to improve the performance. They are:

1) If two successive detected epochs were less than or equal to 15 epochs apart, then all epochs between these two epochs were also assigned as the target class.

2) Because sleep always starts with Awake, if the first detected Awake&REM epoch was not the first epoch in the recording, the epochs from the first epoch to the position of the first detected epoch were all assigned as the target class.

The role of the first rule was to connect epochs detected by the thresholds, because a REM period usually lasts for a long time (30 minutes) [16]. During these long periods, respiration activity would not always be irregular, so the connection scheme will link these periods to irregular periods to form a complete block. Rule 2 allows the system to get started correctly. Fig. 4 gives the comparison of the original detected results and the post processed results. The results demonstrate that by connecting the scattered epochs, most of the REM and Awake stages were detected.

B. Layer 2: Awake vs. REM

The second layer aimed to separate REM and Awake stages. It was hard to separate these two stages because they both have irregular respiration.

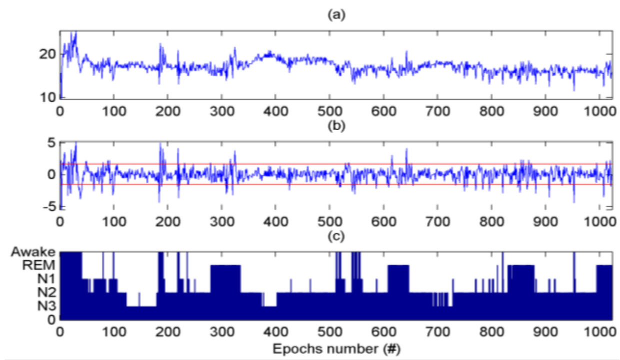

Figure 3. Process of the first layer. (a) The orginal RR, (b) The detrended RR. Two lines indicate two thresholds, (c) ground truth

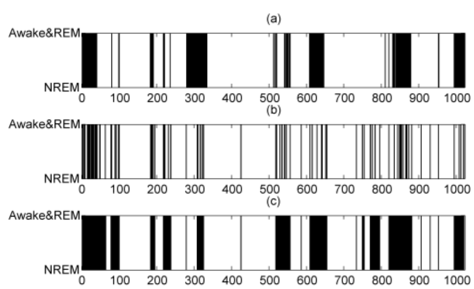

Figure 4. An example of the original detected results and the post processed results. (a) ground truth, (b) detected epochs before post-processing, (c) detected epochs after post-processing.

One assumption is that if a person woke up in the middle of the night suddenly, there might be a more severe fluctuation in respiration, for example, a suddenly shortened breath interval or a deeper breath. This was the reason why the MADI was selected as the feature for this layer. This feature measured the biggest change of the breath intervals in each epoch.

However the feature could only separate part of Awake from REM. So two other rules were applied on the outputs of the first layer in order to improve the results. So there were totally three rules:

1) If the detected Awake&REM epochs in the first layer occur in the first 60 minutes (120 epochs), these epochs were assigned as Awake.

2) If the duration of a block of continuous detected Awake&REM epochs in the first layer was less than 5 minutes (10 epochs), these epochs were assigned as Awake.

3) If the MADI was larger than 4, the epoch was assigned as Awake.

The first two rules were set according to sleep theory: 1) the first REM period usually occurs about 70 minutes after sleep onset [16]; 2) the first REM period is short, but the duration of REM period after that is approximately 30 minutes. The threshold for MADI was picked based on the observation of the data.

IV. RESULTS

For the evaluation of the classification performance, two measures were computed on each recording. They were: accuracy and Cohen's kappa coefficient. The kappa coefficient measures the inter-rater agreement of two observers. Its advantage is a lower sensitivity to an imbalanced dataset. In addition, the sensitivity and specificity were also used to evaluate the performance of the first layer.

The performance measures of the first layer before and after post-processing are shown in Table II. The result shows only 24.45% Awake&REM epochs were successfully detected without post-processing. After post-processing, the sensitivity increased about 50% while only losing 16% of the specificity. The post-processing was thus shown to be efficient.

The three class confusion matrix of total 21 recordings is displayed in table III. Accuracy is 74.00%±5.30% and Kappa is 0.49±0.08. In addition, the sleep efficiency (S_{eff}) was estimated using the results of the classification and compared to the true values. Sleep efficiency is defined as:

$$S_{eff} = \text{total sleep time} / \text{time in bed} \tag{4}$$

TABLE II. PERFORMANCE MESURES OF THE FIRST LAYER BEFORE AND AFTER POST-PROCESSING

	Sensitivity (%)	**Specificity (%)**	**Accuracy (%)**	**kappa**
Before	24.45±3.36	94.06±1.96	72.79±4.91	0.22±0.05
After	76.08±9.83	78.05±6.94	77.05±5.30	0.49±0.1

TABLE III. CONFUSION MATRIX OF ALL RECORDINGS

Actual / Output	**NREM**	**REM**	**Awake**
NREM	10826	1065	408
REM	2405	2960	521
Awake	740	83	1030

TABLE IV. COMPARISON OF PERFORMANCE WITH PREVIOUS WORKS

Author/year	**Features**	**Accuracy (%)**	**kappa**
Long, 2014[9]	27 RV	77.10±7.60	0.48±0.17
Xiao, 2013[10]	41 HRV	72.58±6.70	0.46±0.10
Mendez, 2010 [11]	17 HRV	71.95±7.47	0.42±0.10
Proposed method	2 RV	74.00±5.30	0.49±0.08

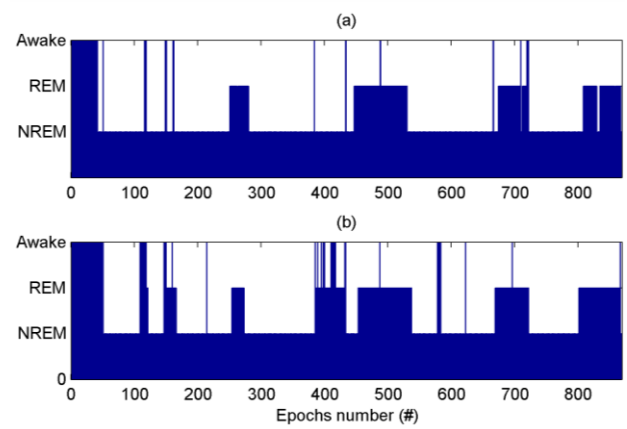

Figure 5. An example outputs of one recording (k=0.67). (a) Ground truth, (b) output

The sleep efficiency thus estimated had 3.61% ±3.66% absolute error for all 21 recordings.

Fig.5 displays the outputs of one recording with the kappa coefficient of 0.67. Except the epochs around 400, most of other epochs were assigned to the right class with minor errors.

Table IV shows the comparison between performance of the proposed method and previous work using different data.

V. Discussion

Our proposed method shows state-of-the-art result performance for Awake/REM/NREM classification. Most importantly only two features and the simple threshold comparison classifier was used. While we tested more sophisticated classifiers such as Support Vector Machines and Random Forests, we found that these features could be used in a simple and fast rule based system to obtain better classification results. Hence, we opted for this simple approach.

The detrended RR feature successfully detected rough positions of Awake and REM periods. By combining with two additional rules as post-processing, sensitivity increased from 24.45% to 76.08%. It demonstrated that commonsense rules, based on sleep theory, do improve results when Awake and REM are placed in the same class.

The Awake and REM separation in the second layer is a hard problem, because these two stages have similar physiological indexes including respiratory activities. The results also showed such situation. The MADI feature and two additional rules only separated about half of Awake from REM in the second layer. However, the error of sleep efficiency of 3.61% was promising. Sleep efficiency is one of the key measures to evaluate sleep quality. Thus, the proposed method can be used to reflect sleep quality.

From the presented results, we conclude that the proposed algorithm is efficient for sleep stage recognition. However, because sleep can be very different among individuals and different sensors also may introduce various types of noise, more experiments are needed to evaluate the robustness of the approach. Our next step will be collecting data using our bed sensor in order to achieve the goal of long-term non-invasive sleep monitoring. Furthermore, by combining HRV and body movement features extracted from the bed sensor, the classification results can be improved over current results, especially the detection rate of the Awake stage.

VI. Conclusion

A three stage (Awake, REM and NREM) classification system was implemented with 21 recordings from the Sleep-EDF database. A two-layer classification system was developed with only two RV features and the threshold comparison classifier. Some rules were added based on general sleep theories. The method achieved the accuracy of 74.00% ±5.30% and Kappa of 0.49 ±0.08. The error of sleep efficiency was 3.61% ±3.66%. This approach is comparable with previous work using simpler features. We conclude the proposed method has great potential for application to the bed sensor data in the future.

References

[1] G. Garcia-Molina, F. Abtahi, M. Lagares-Lemos, "Automated NREM sleep staging using the Electro-oculogram: A pilot study," in *Proc. 34th Annu. Int. Conf. IEEE Eng. in Medicine and Biological Soc.*, San Diego, CA, 2012, pp.2255-2258.

[2] I. H. Hansen, M. Marcussen, J. A. E Christensen, P. Jennum, H. B. D. Sorensen, "Detection of a sleep disorder predicting Parkinson's disease," in *Proc. 35th Annu. Int. Conf. IEEE Eng. in Medicine and Biological Soc.*, Osaka, Japan, 2013, pp.5793-5796.

[3] *Alzheimer's Disease and Sleep* [Online]. Available: https://sleepfoundation.org/sleep-disorders-problems/alzheimers-disease-and-sleep.

[4] A. Rechtschaffen, A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," U.S. Dept. of Health, Education, and Welfare, 1968.

[5] M. H. Silber et al., "The visual scoring of sleep in adults," *J. Clinical Sleep Medicine*, vol. 3, no. 2, pp. 121-131, Mar. 2007.

[6] American Academy of Sleep Medicine and C. Iber, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications," American Academy of Sleep Medicine, 2007.

[7] B. Yilmaz, M. H. Asyali, E. Arikan, S. Yetkin, F. Özgen, "Sleep stage and obstructive apneaic epoch classification using single-lead ECG," *BioMedical Engineering Online*, Vol. 9, no.2, pp. 39-39, Aug. 2010.

[8] D. Heise and M. Skubic, "Monitoring pulse and respiration with a hydraulic bed sensor," in *Proc. 32rd Annu. Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, Buenos Aires, Argentina, 2010, pp. 1926–1929.

[9] X. Long, J. Yang, T. Weysen, R. Haakma, J. Foussier, P. Foneseca, and R. M. Aarts, "Measuring dissimilarity between respiratory effort signals based on uniform scaling for sleep staging," *Physiological Measurement*, vol. 35, pp. 2529–2542, Nov. 2014.

[10] M. Xiao, H. Yan, J. Song, Y. Yang, and X. Yang, "Sleep stages classification based on heart rate variability and random forest," *Biomedical Signal Process. Control*, vol. 8, pp. 624–633, 2013.

[11] M. O. Mendez, M. Migliorini, J. M. Kortelainen, D. Nistico, E. Arce-Santana, S. Cerutti, A. M. Bianchi, "Evaluation of the sleep quality based on bed sensor signals: Time-variant analysis," in *Proc. 32rd Annu. Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, Buenos Aires, Argentina, 2010, pp. 3994-3997.

[12] A. L. Goldberger AL et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, pp. e215-e220, Jun. 2000.

[13] B Kemp, "Model-based monitoring of human sleep stages," Ph.D. dissertation, Twente Univ., Enschede, Netherlands, 1987.

[14] K. E. Barett, S. M. Barman, S. Boitano, H. L. Brooks, "Respiratory Physiology," in *Ganong's review of medical physiology*, 24th ed. New York: McGraw-Hill Medical, 2012, sec. VI, pp. 619.

[15] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Amer. stat. assoc.*, vol. 74, pp. 829-836, Dec. 1979.

[16] R. W. McCarley, "Neurobiology of REM and NREM sleep," *Sleep Medicine*, vol. 8, pp. 302-330, Jun. 2007.