Estimation of A Priori Probabilities of Sleep Stages: A Cycle-based Approach

Alexander Tataraidze, *Member, IEEE*, Lesya Anishchenko, Lyudmila Korostovtseva
Mikhail Bochkarev, Yurii Sviryaev and Sergey Ivashov

***Abstract*— This paper presents a model for the estimation of a priori probabilities of sleep epoch classes based on the epoch location in a sleep cycle. These probabilities are used as additional features for sleep stage classification based on the analysis of respiratory effort. The model was validated with data of 685 subjects selected from the Sleep Heart Health Study dataset. The model improves a base algorithm by 8 percent points and demonstrates Cohen's kappa of 0.56 ± 0.12. Our results will contribute to the development of screening tools for unobtrusive sleep structure estimation.**

I. Introduction

Sleep structure is a key parameter for the diagnosis of sleep disorders. The gold standard for sleep structure estimation is polysomnography (PSG). A skilled expert — a physician or a technician — classifies every epoch (30 s. interval of signals) based on the visual analysis of PSG signals, namely electroencephalogram (EEG), electromyogram (EOG) and electrooculogram (EMG). An epoch might be classified as one of the following classes according to Rechtschaffen and Kales (R&K) rules: wakefulness (W), rapid eye movement sleep (R, REM), first (N1), second (N2), third (N3) or fourth (N4) stage of non-REM sleep (NREM). Sometimes, N3&N4 are called deep sleep (D), and N1&N2 are merged into the light sleep (L) class.

PSG is an inconvenient, labour-consuming and expensive procedure. Therefore, the topic of unobtrusive sleep monitoring became popular in the recent decade. Most studies are focused on sleep stage classification based on the analysis of cardiorespiratory features. Some of them also used a priori probabilities of epoch class to improve classification results.

A priori probabilities, as we will understand it here, are probabilities of epoch class which were estimated by the analysis of the epoch location without using features extracted from signals. This technique is based on the simple idea that the probabilities depend on epoch location. E.g. we can expect higher probability of wakefulness at the beginning and at the end of a record, REM epochs are more common in the second part of sleep, while deep sleep epochs are more frequent in the first. Fig. 1 shows time-varying a priori probabilities throughout a whole record.

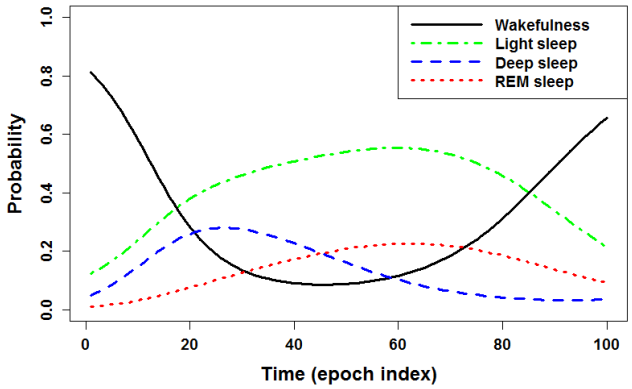

Fig. 1. Time-varying a priori probabilities of sleep stages. The probabilities were estimated based on the relative frequency of sleep stages for epochs with the same index in dataset described in Table I. The epoch index is between 1 and 100.

Based on this idea we used a normalized epoch index as one of the features for the epoch classification in our previous study [1]. In its turn, Redmond et al. [2] and Long et al. [3] used time-varying a priori probabilities as a parameter of a discriminant classifier. The probabilities were estimated based on the relative frequency of classes for epochs with the same index.

Sleep consists of NREM-REM cycles. Normal sleep includes 4–6 cycles. The typical length of the cycle is 70–120 minutes. So, it should be possible to estimate a priori probabilities more accurately if taking into consideration an epoch location in a cycle rather than in a record.

The aim of the study is to verify the hypothesis that a priori probabilities estimation based on epoch location in a sleep cycle might improve sleep stage classification in comparison to the base algorithm.

II. Materials and Methods

The diagram of the proposed method is presented on Fig. 2. The method includes four main stages. The first stage is the base algorithm for sleep stage classification based on the analysis of the respiratory inductive plethysmography signal (RIP). The second stage is splitting record to sleep cycles. The third stage is estimation of a priori probabilities based on epoch location in a sleep cycle. The fourth stage is final epoch classification.

*Research supported by the grant of Russian Foundation for Basic Research (15-07-02472 A)

A. Tataraidze, L. Anishchenko and S. Ivashov are with Bauman Moscow State Technical University. Russian Federation,105005, Moscow, 2nd Baumanskaya str., 5 (e-mail: tataraidze@rslab.ru, tel.: +7(495)632-22-19).

L. Korostovtseva, M. Bochkarev and Y. Sviryaev are with V.A. Almazov Federal North-West Medical Research Center. St. Petersburg, Russian Federation.

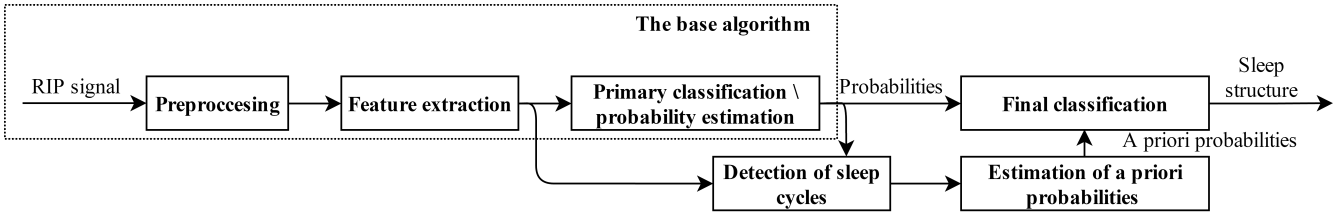

Fig. 2. The diagram of the proposed method

A. Dataset

We used PSG records from the open PSG database of the Sleep Heart Health Study (SHHS) [4]–[6]. The database consists of two parts: SHHS1 (N = 5793) and SHHS2 (N = 2651), which were collected from the same subjects during 1995–1998 (SHHS1) and 2001–2003 (SHHS2) years.

Data from both parts were used to compute coefficients of the model, which used for the estimation of a priori probabilities. Only records with Apnea-Hypopnea Index (AHI) $\leq$ 5 episodes/h, which is normal value, were used in the study. In total 1636 records were involved in the development of the model (Table I).

Data from SHHS1 only were used to train classifiers and validate the method to prevent using data of the same subjects in training and test sets. The following exclusion criteria were applied: AHI $\geq$ 5 episodes/h; hours of applicable thoracic RIP signal is less than time in bed by at least one hour. Moreover, 3 records were removed due to technical reasons. In total 685 records were used to train classifiers and validate the method (Table II).

AHI was calculated based on episodes of apneas and hypopneas with $\geq$ 3% oxygen desaturation or arousals according to the AASM rules. Hours of applicable signals were evaluated by trained technicians and provided with SHHS database, as well as time in bed. PSG records were scored according to the R&K rules.

B. The base algorithm

1) Data preprocessing: Thoracic RIP signal was used for the feature extraction. The signal was filtered with a Butterworth low-pass filter at cut-off frequency of 0.6 Hz. Peaks and troughs were detected based on the search of turning points.

"Wake/Movement" and "Unscored" epoch labels were changed to W. N1 and N2 epochs were combined to L class. N3 and N4 epochs were combined to D class. Thus, each epoch might be belonged to one of the following classes: W, R, L, D.

2) Feature extraction: We used features described in our previous paper [7], except normalized epoch index, the length and the energy of artifact periods.

3) Classification: Gradient boosted trees from XGBoost library were used for the classification purpose [8]. The classifier predicts epoch class probabilities.

C. Splitting record to sleep cycles

The end of the cycle is the end of the REM stage. For the REM/~~REM~~ binary classification we used XGBoost classifier, features extracted from the RIP signal, and the probabilities from the base algorithm as additional features. ~~REM~~ is class uniting all non-REM epochs including W. The end of cycle was determined as end of REM epoch if at least 5 of 8 epochs from the left are REM and 10 epochs from the right are ~~REM~~.

D. Model

The mathematical model for a priori probabilities estimation is a system of logistic regression equations. Let $P_m = \{p_W, p_L, p_D, p_R\}$ be a priori probabilities of epoch to be belonged to $\{W, L, D, R\}$ classes estimated by the model:

$$\begin{cases} p_W = \dfrac{1}{1+e^{-y_W}} \\ p_R = \dfrac{1}{1+e^{-y_R}} \\ p_L = \dfrac{1}{1+e^{-y_L}} \\ p_D = \dfrac{1}{1+e^{-y_D}} \end{cases}$$

Since we can expect that the first and the last cycles would be different from intermediate cycles, let define $y_i, i \in \{W, L, D, R\}$, by different equation:

TABLE I. The Dataset for Model Development (N = 1636)

Total Number of Epochs	1760763
SHHS1	1130
SHHS2	506
Male:Female	404:1232
Age (years)	60.77 ± 11.51(39.00 – 90.00)
Sleep Efficiency (%)	76.19 ± 17.27(36.54 – 99.05)
Wakefulness (%)	29.38 ± 12.42(2.62 – 82.99)
REM (%)	14.58 ± 5.90(0.00 – 35.73)
Light sleep (%)	42.31 ± 11.92(6.48 – 93.64)
Deep sleep (%)	13.73 ± 8.83(0.00 – 51.38)

Mean ± SD (range)

TABLE II. The Dataset for Training Classifiers and Validation of the Method (N = 685)

Total Number of Epochs	693433
Male:Female	177:508
Age (years)	59.31 ± 11.99 (39 - 90)
Body Mass Index (kg/m^2)	25.54 ± 4.05 (18.00 - 50.00)
Apnea Hypopnea Index (episodes/h.)	3.03 ± 1.30 (0.16 - 4.99)
Sleep Efficiency (%)	85.75 ± 9.00 (36.19 - 98.53)
Wakefulness (%)	26.50 ± 11.41(4.38 - 72.62)
REM (%)	14.84 ± 5.85 (0.00 - 29.34)
Light sleep (%)	43.99 ± 11.76 (6.48 - 91.33)
Deep sleep (%)	14.68 ± 9.31 (0.00 - 45.44)

Mean ± SD (range)

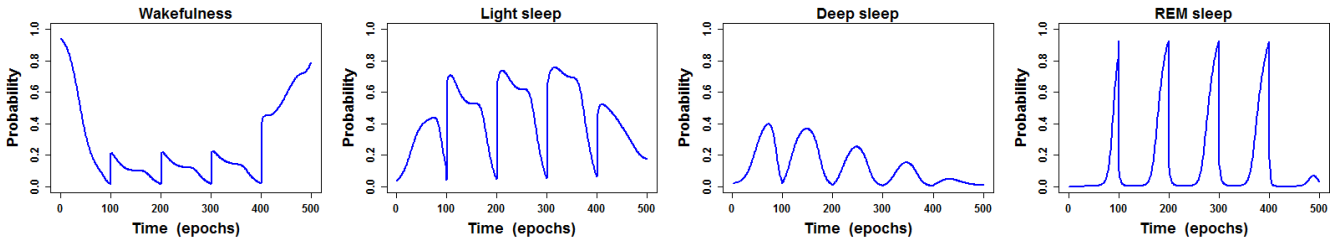

Fig. 3. A priori probabilities of sleep stages estimated by the model based on an epoch position in a sleep cycle. The record consists of 5 100-epochs cycles.

$$y_i = \begin{cases} \beta_0 + \beta_1 t^3 + \beta_2 t^2 + \beta_3 t & c = 1 \\ \beta_0 + \beta_1 t^3 + \beta_2 t^2 + \beta_3 t + \beta_4 c & 1 < c < N \\ \beta_0 + \beta_1 t^3 + \beta_2 t^2 + \beta_3 t & c = N \end{cases}, \quad (1)$$

where t — normalized epoch index, c — cycle index, N — total number of cycles in the record. The sum of the a priori probabilities must not exceed 1, so we need to normalize them:

$$p_i = \frac{p_i}{\sum_{i \in \{W,L,D,R\}} p_i}.$$

The β coefficients were computed based on data presented in Table I. If it is not possible to split the record to cycles, a priori probabilities $\{p_W, p_L, p_D, p_R\}$ for all epoch in the records were set to constants $\{0.29, 0.42, 0.14, 0.15\}$, which correspond to class frequencies in the dataset (Table I).

E. Final classification

The final classification was produced by a simple neural network with single hidden layer. The layer consists of 4 units. Each epoch was described by 8 features — 4 probabilities from the primary classifier and 4 a priori probabilities from the model.

TABLE III. Model Coefficients for the First Cycle

	Value	Std. error	Std. coefficient	Z-value
β_0^W	3.5	2.4×10^{-2}		148.8
β_1^W	-8.0×10^{-6}	2.7×10^{-7}	−4.7	−29.2
β_2^W	1.4×10^{-3}	4.1×10^{-5}	8.3	33.2
β_3^W	-1.3×10^{-1}	1.8×10^{-3}	−7.6	−72.3
β_0^L	−3.3	2.3×10^{-2}		−143.3
β_1^L	-5.3×10^{-6}	2.1×10^{-7}	−3.4	−25.9
β_2^L	-5.8×10^{-5}	3.3×10^{-5}	−0.4	−1.7*
β_3^L	7.4×10^{-2}	1.6×10^{-3}	4.7	46.2
β_0^D	−3.8	3.3×10^{-2}		−114.5
β_1^D	-2.2×10^{-5}	2.8×10^{-7}	−16.0	−78.3
β_2^D	2.2×10^{-3}	4.5×10^{-5}	16.9	49.11
β_3^D	1.4×10^{-3}	2.2×10^{-3}	0.1	0.6*
β_0^R	−8.9	2.9×10^{-1}		−31.1
β_1^R	1.2×10^{-5}	1.1×10^{-6}	11.4	11.3
β_2^R	-7.9×10^{-4}	2.2×10^{-4}	−7.6	−3.6
β_3^R	6.9×10^{-2}	1.4×10^{-2}	6.3	4.9

* — the predictor is non-significant according to Wald test. Other predictors are statistically significant, p < 0.001.

F. Experiments

The validation of the proposed method was conducted by 5-fold cross-validation. The dataset described in Table II was divided into 5 subsets, each of them consisted of 137 subjects. A training set was formed from 4 subsets and data of the last remaining subset was used as a test set. That was repeated 5 times with changing subset included in the test set. Cohen's kappa (κ) coefficient was computed for each subject from test set. The Cohen's kappa is a coefficient of inter-rater agreement, which takes into consideration the possibility of the agreement occurring accidentally. The results for each subject were collected together, the mean and the standard deviation were calculated. The metric were computed for the base algorithm and the entire method with the model.

III. Results

Values and standard errors of model coefficients, standardized coefficients and Z-values are presented in Tables III–V. The model output for the record consisted of 5 100-epochs cycles is presented on Fig. 3.

The mean κ of 0.56 were achieved using the model, which improves the base algorithm by 8 percent points. Table VII shows the results of the algorithm modification.

TABLE IV. Model Coefficients for Intermediate Cycles

	Value	Std. error	Std. coefficient	Z-value
β_0^W	−1.2	1.3×10^{-2}		−94.9
β_1^W	-9.5×10^{-6}	2.0×10^{-7}	−8.6	−46.3
β_2^W	1.2×10^{-3}	2.9×10^{-5}	11.1	40.7
β_3^W	-5.3×10^{-2}	1.1×10^{-3}	−4.8	46.9
β_4^W	-1.1×10^{-1}	3.3×10^{-3}	0.3	33.9
β_0^L	0.8	1.1×10^{-2}		71.8
β_1^L	-1.1×10^{-5}	1.4×10^{-7}	−6.3	−78.6
β_2^L	1.1×10^{-3}	2.0×10^{-5}	6.4	52.4
β_3^L	-3.8×10^{-2}	8.7×10^{-3}	−2.2	−44.2
β_4^L	-1.6×10^{-1}	2.3×10^{-3}	0.3	69.9
β_0^D	−2.7	2.5×10^{-2}		−121.9
β_1^D	-2.8×10^{-7}	2.6×10^{-7}	−0.2	−1.1*
β_2^D	-1.4×10^{-3}	3.8×10^{-5}	−12.2	−37.7
β_3^D	1.3×10^{-1}	1.6×10^{-3}	10.9	80.4
β_4^D	-7.0×10^{-1}	4.1×10^{-3}	−1.9	−171.4
β_0^R	−1.9	2.1×10^{-2}		−93.0
β_1^R	-4.0×10^{-5}	3.2×10^{-7}	−28.1	−124.1
β_2^R	8.0×10^{-3}	5.4×10^{-5}	58.4	148.3
β_3^R	-3.6×10^{-1}	2.3×10^{-3}	−25.2	−150.9
β_4^R	2.5×10^{-1}	3.9×10^{-3}	0.6	64.2

* — the predictor is non-significant according to Wald test. Other predictors are statistically significant, p < 0.001.

TABLE V. Model Coefficients for the Last Cycle

	Value	Std. error	Std. coefficient	Z-value
β_0^W	-2.5×10^{-1}	1.4×10^{-2}		-17.7
β_1^W	-2.8×10^{-6}	1.9×10^{-7}	-1.6	-15.0
β_2^W	4.7×10^{-4}	2.9×10^{-5}	2.9	16.4
β_3^W	-4.5×10^{-3}	1.2×10^{-3}	-0.3	-3.6
β_0^L	-6.4×10^{-2}	1.4×10^{-2}		-4.4
β_1^L	3.4×10^{-6}	2.0×10^{-7}	2.0	17.4
β_2^L	-6.5×10^{-4}	2.9×10^{-5}	-4.1	-22.0
β_3^L	1.5×10^{-2}	1.2×10^{-3}	0.9	12.3
β_0^D	-4.6	5.3×10^{-2}		-86.4
β_1^D	1.2×10^{-5}	6.6×10^{-7}	17.9	17.9
β_2^D	-2.3×10^{-3}	9.9×10^{-5}	-23.4	-23.4
β_3^D	1.1×10^{-1}	4.3×10^{-3}	26.2	26.2
β_0^R	-8.0×10^{-1}	3.8×10^{-2}		-21.3
β_1^R	-7.4×10^{-5}	1.3×10^{-6}	-153.4	-59.2
β_2^R	1.3×10^{-2}	2.1×10^{-4}	288.3	62.5
β_3^R	-6.2×10^{-1}	9.4×10^{-3}	-128.2	-66.0

All predictors are statistically significant, $p < 0.001$.

The comparison of our results with state-of-the-art are presented in Table VI.

TABLE VI. Our Results vs State-of-the-Art for WRLD Classification Based on Cardiorespiratory Parameters

First author, year	Signals	N of subj.	κ
Hedner, 2011 [12]	PAT, PO, ACT	227	0.48
Willemen, 2014 [13]	ECG, RIP, ACT	36	0.56
Long, 2014 [14]	RIP	48	0.41 ± 0.14
Fonseca, 2015 [15]	ECG, RIP	48	0.49 ± 0.13
Tataraidze, 2016 [1]	ECG, RIP	625	0.57 ± 0.13
Our study	**RIP**	**685**	0.56 ± 0.12

ECG — electrocardiography, RIP — respiratory inductance plethysmography, ACT — actigraphy, PAT — peripheral arterial tone, PO — pulse oximetry. Mean ± SD.

TABLE VII. Classification Performance Change

Modification	Cohen's kappa
The base algorithm	0.4781 ± 0.1292
+ The model	0.5579 ± 0.1196

There is a significant difference between the results according to Wilcoxon signed-rank test ($p < 0.001$). Mean ± SD.

IV. Discussion

The proposed method showed a good performance. It was tested on a large population, which confirms its reliability. Our study proved the hypothesis that a priori probabilities estimation based on epoch location in a sleep cycle might improve sleep stage classification. To the best of our knowledge, this is the first study which used a priori probabilities estimated based on cycles. The performance of proposed method is competitive with those which used combination of respiratory and cardio analysis (Table VI).

The method shifts subject's sleep structure to the average, so it might be not applicable for subjects with sleep disorders. It is a relevant topic for further research.

Sleep changes with the age — seniors have shorter REM and deep sleep as well as total sleep time [9]. Furthermore, sleep in men and women is different [10], [11]. Thus, the future improvement of the model might be done by personalization of a priori probability estimation. It can be done by adding new parameters to the equation (1) such as age, gender, etc.

The model might be used not only with RIP but with other unobtrusive methods, such as electrocardiography, ballistocardiography, bioradiolocation, etc. Apparently, the model is not applicable for PSG signals (EEG, EOG, EMG) which is highly informative itself. The results of the study will contribute to the development of screening tools for unobtrusive sleep structure estimation.

References

[1] A. Tataraidze, L. Korostovtseva, L. Anishchenko *et al.*, "Sleep Architecture Measurement Based on Cardiorespiratory Parameters," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 53, Orlando, USA, 2016, pp. 3478–3481.

[2] S. J. Redmond, P. De Chazal, C. O'Brien *et al.*, "Sleep staging using cardiorespiratory signals," *Somnologie*, vol. 11, no. 4, pp. 245–256, 2007.

[3] X. Long, J. Foussier, P. Fonseca *et al.*, "Analyzing respiratory effort amplitude for automated sleep stage classification," *Biomedical Signal Processing and Control*, vol. 14, no. 1, pp. 197–205, 2014.

[4] S. Redline, M. H. Sanders, B. K. Lind *et al.*, "Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group." *Sleep*, vol. 21. no. 7. pp. 759–67, 1998.

[5] D. Dean, A. Goldberger, R. Mueller *et al.*, "Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource," *Sleep*, vol. 39, no. 5, pp. 1151–1164, 2016.

[6] S. F. Quan, B. V. Howard, C. Iber *et al.*, "The Sleep Heart Health Study: Design, Rationale, and Methods," *Sleep*, vol. 2, no. 12, pp. 1077–1085, 1997.

[7] A. Tataraidze, L. Korostovtseva, L. Anishchenko *et al.*, "Bioradiolocation-based Sleep Stage Classification," in *Annual International Conference of the IEEE Engineering in Medicine and Biology* , vol. 38, Orlando, USA, 2016, pp. 2839–2842.

[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *arxiv.org*, 2016.

[9] K. L. Lichstein, H. H. Durrence, B. W. Riedel *et al.*, *Epidemiology of Sleep: Age, Gender, and Ethnicity*. Mahwah, New Jersey: Lawewnce Erlbaum Assiciates, 2004.

[10] K. I. Hume, F. Van, and A. Watson, "A field study of age and gender differences in habitual adult sleep." *Journal of Sleep Research*, vol. 7, no. 2, pp. 85–94, 1998.

[11] A. Silva, M. L. Andersen, M. T. De Mello *et al.*, "Gender and age differences in polysomnography findings and sleep complaints of patients referred to a sleep laboratory," *Brazilian Journal of Medical and Biological Research*, vol. 41, no. 12, pp. 1067–1075, 2008.

[12] J. Hedner, D. P. White, A. Malhotra *et al.*, "Sleep Staging Based on Autonomic Signals : A Multi-Center Validation Study," *Journal of Clinical Sleep Medicine*, vol. 7, no. 3, pp. 301–306, 2011.

[13] T. Willemen, D. Van Deun, V. Verhaert *et al.*, "An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 661–669, 2014.

[14] X. Long, J. Yang, T. Weysen *et al.*, "Measuring dissimilarity between respiratory effort signals based on uniform scaling for sleep staging," *Physiological Measurement*, vol. 35, no. 12, pp. 2529–2542, 2014.

[15] P. Fonseca, X. Long, M. Radha *et al.*, "Sleep stage classification with ECG and respiratory effort," *Physiological Measurement*, vol. 36, no. 10, pp. 2027–2040, 2015.