Improving Sleep/Wake Detection via Boundary Adaptation for Respiratory Spectral Features

Xi Long[1], *Member, IEEE*, Reinder Haakma[2], Jérôme Rolink[3], *Member, IEEE*, Pedro Fonseca[1], and Ronald M. Aarts[1], *Fellow, IEEE*

***Abstract*— In previous work, respiratory spectral features have been successfully used for sleep/wake detection. They are usually extracted from several frequency bands. However, these traditional bands with fixed frequency boundaries might not be the most appropriate to optimize the sleep and wake separation. This is caused by the between-subject variability in physiology, or more specifically, in respiration during sleep. Since the optimal boundaries may relate to mean respiratory frequency over the entire night. Therefore, we propose to adapt these boundaries for each subject in terms of his/her mean respiratory frequency. The adaptive boundaries were considered as those being able to maximize the separation between sleep and wake states by means of their mean power spectral density (PSD) curves overnight. Linear regression models were used to address the association between the adaptive boundaries and mean respiratory frequency based on training data. This was then in turn used to estimate the adaptive boundaries of each test subject. Experiments were conducted on the data from 15 healthy subjects using a linear discriminant classifier with a leave-one-subject-out cross-validation. We reveal that the spectral boundary adaptation can help improve the performance of sleep/wake detection when actigraphy is absent.**

I. Introduction

Objective assessment of sleep is often achieved by monitoring sleep and wake phases during bedtime. Polysomnography (PSG) recordings with manually scored hypnograms (done by sleep technicians with a 30-sec epoch basis) are the gold standard for objective sleep analysis [1], [2]. They are typically acquired in sleep laboratories.

Human respiratory activity during sleep has been proved to associate with autonomic nervous system and breathing control [3], [4]. It has been increasingly used for sleep stage classification in recent years [5], [6], [7] as long as it can be unobtrusively obtained with e.g., textile bed sensors [8], Doppler radar [9], and camera [10]. For that purpose, many respiratory features have been investigated. For example, features extracted based on the power spectral density (PSD) analysis of respiratory signal have been shown to be effective in classifying sleep and wake states [5], [11]. They are the logarithms of spectral powers computed in different frequency bands including a very low frequency band (VLF: 0.01-0.05 Hz), a low frequency band (LF: 0.05-0.15 Hz), and a high frequency band (HF: 0.15-0.5 Hz) and the ratio of spectral powers between LF and HF bands. These features are assumed to express different aspects of respiratory dynamics, although they were not well-defined in physiology. We speculate that HF power could associate with the level of dominance in respiratory frequency, LF power might associate with short-range correlations or breathing control, and VLF power might relate to long-range fluctuations [12]. In general, a more spread distribution in respiratory PSD should indicate a less regular breathing rhythm.

Yet we observed large variations on the feature values during sleep (or wake) state between subjects. It may be, partially, caused by the presence of between-subject difference in physiology. This difference is expected to correspond to the mean respiratory frequency over epochs throughout the night. Since the boundaries of the spectral bands were fixed when computing the features, they might not be optimized for separating sleep and wake states for each subject. We therefore propose to adapt these spectral boundaries for each subject by modeling the association between the "optimal" boundaries and the mean respiratory frequency based on the data from other subjects. This is expected to reduce the between-subject variation to some extent, and thus help improve sleep/wake detection with respiratory signals. Here, the respiratory frequency of each epoch can be computed as the the dominant frequency within the LF and HF bands.

To evaluate the capability of the respiratory spectral features in separating sleep and wake states, the Mahalanobis distance [13] was utilized. We adopted a linear discriminat (LD) classifier, which has been widely used in sleep/wake detection [5], [6], [11], [14]. This preliminary study examined whether our proposed method can help improve the results in classifying sleep and wake states.

II. Materials and Methods

A. Data Set

A date set comprising full PSG and actigraphy (Actiwatch, Philips Respironics) of fifteen healthy subjects (ten females, age 31.0 $\pm$ 10.4 years) was used. Nine subjects were monitored (Alice 5 PSG, Philips Respironics) at the Sleep Health Center, Boston, USA during 2009 and six were monitored using (Vitaport 3 PSG, TEMEC) in the Philips Experience Lab, Eindhoven, the Netherlands during 2010. The subjects had a Pittsburgh Sleep Quality Index (PSQI) of less than 6. Respiratory effort was measured with respiratory inductance plethysmography (sampled at 10 Hz). According to the American Academy of Sleep Medicine guidelines [2],

[1]X. Long, P. Fonseca, and R. M. Aarts are with Department of Electrical Engineering, Eindhoven University of Technology, Postbox 513, 5600 MB, Eindhoven, The Netherlands and with the Philips Research, HTC34, 5656 AE, Eindhoven, The Netherlands `xi.long.ee@gmail.com`

[2]R. Haakma is with the Philips Research, HTC34, 5656 AE, Eindhoven, The Netherlands `reinder.haakma@philips.com`

[3]J. Rolink is with the Chair of Medical Information Technology, RWTH Aachen University, Pauwelsstrasse 20, 52074 Aachen, Germany `rolink@hia.rwth-aachen.de`

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

manual scoring of 30-sec epochs was performed by a sleep technician and then each epoch was labeled as *sleep* if it was scored as rapid-eye-movement (REM), N1, N2, or N3 sleep, otherwise as *wake*. The average total recording time was 7.2 ± 1.1 hours and the sleep efficiency was 92.3 ± 3.8%.

B. Respiratory Spectral Features

The respiratory effort signals were preprocessed before they were used to extract features. They were first low-pass filtered (10th-order Butterworth filter with a cutoff frequency of 0.6 Hz) to eliminate high frequency noise. Afterwards, they were then normalized by subtracting the median peak-to-trough amplitude estimated over the entire recording to remove the signal baseline [5], [7].

Since the respiratory spectral features were computed in the frequency domain of the respiratory effort signal, we estimated the power spectral density (PSD) for each epoch using a fast Fourier transform with a Hanning window based on the resulting preprocessed signal. As mentioned, we considered four respiratory spectral features: the logarithms of the spectral powers within the VLF, LF, and HF bands and the ratio of LF and HF spectral powers [5]. Note that the PSD of each band was normalized by dividing it by the total the total spectral power in the LF and HF bands.

C. Adaptive Spectral Boundaries

Figure 1 illustrates an example of the averaged and normalized PSD of sleep and wake epochs from a subject. It seems that the "optimal" frequencies for separating the sleep and wake states occur at around 0.22 Hz and 0.32 Hz, which do not equal the traditional fixed boundaries. This means that using the fixed boundaries to compute the respiratory spectral features would result in errors when classifying sleep and wake epochs. To tackle this problem, we proposed to use subject-specific adaptive boundaries to designate new frequency bands (VLF⋆, LF⋆, and HF⋆) instead of using the fixed boundaries. As we hypothesized, these boundaries might relates to mean respiratory frequency ω over all epochs throughout the night regardless of their state (sleep or wake).

Let B_x^u and B_x^l express the adaptive upper and lower boundaries of the x band (VLF⋆, LF⋆, or HF⋆), respectively. We have $B_{\mathrm{HF}}^l = B_{\mathrm{LF}}^u$ and $B_{\mathrm{LF}}^l = B_{\mathrm{VLF}}^u$. The lower and upper boundaries of the VLF⋆ band are set to be fixed (B_{VLF}^l = 0.01 Hz and B_{VLF}^u = 0.05 Hz). Hence, the problem is addressed to locate B_{HF}^l (or B_{LF}^u) and B_{HF}^u for each subject. As illustrated in Fig.1, we consider the optimal boundaries (B_{HF}^l and B_{HF}^u) of the LF⋆ and HF⋆ bands that occur at the frequencies corresponding to the 'crossing points' between the average (normalized) PSD curves of sleep and wake states. This should enhance the difference of the area under the PSD curve between the two states in the LF⋆ or HF⋆ band. To obtain the mean respiratory frequency over the entire night, we first considered the respiratory frequency as the frequency of the PSD peak in the traditional LF and HF bands (0.01-0.5 Hz) for each epoch. This is because the respiratory frequency of healthy people usually lies within this range [3]. Afterwards, the mean respiratory frequency ω was calculated as the mean of those peak frequencies over all epochs throughout the whole night.

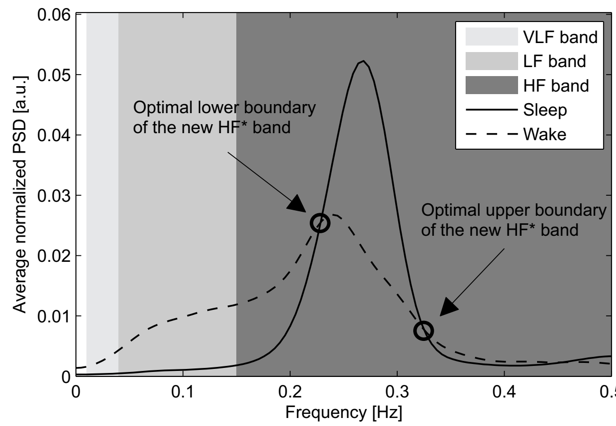

Fig. 1. An example of the average (normalized) PSD of sleep and wake epochs from a subject. The traditional frequency bands are filled and the optimal lower (B_{HF}^l) and upper (B_{HF}^u) boundaries of the new HF⋆ band in separating sleep and wake epochs are indicated with arrows.

To model the association between the optimal boundaries (for B_{HF}^l and B_{HF}^u) and the mean respiratory frequency, a linear regression [15] was simply used

$$\vec{B} = \alpha \cdot \vec{\omega} + \beta + \epsilon, \tag{1}$$

where $\vec{B} = \{B_1, B_2, \ldots, B_i, \ldots, B_n\}$ $(i = 1, 2, \ldots, n)$ is a set of optimal boundaries (B_{HF}^l or B_{HF}^u) from different subjects, $\vec{\omega} = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_n\}$ $(i = 1, 2, \ldots, n)$ are the corresponding mean respiratory frequencies from the n subjects, α and β are regression coefficients (i.e., slope and intercept), and ϵ is an error term. We determined the two coefficients using a least squares estimation (LSE) method [14] such that

$$\alpha = \frac{n \sum_{i=1}^n \omega_i B_i - (\sum_{i=1}^n \omega_i \sum_{i=1}^n B_i)}{n \sum_{i=1}^n \omega_i^2 - (\sum_{i=1}^n \omega_i)^2}, \tag{2}$$

$$\beta = \frac{1}{n}(\sum_{i=1}^n B_i - \alpha \sum_{i=1}^n \omega_i). \tag{3}$$

Then the optimal lower and upper HF boundaries for a new subject k can be estimated by

$$B_{k,\mathrm{HF}}^l = \alpha_{\mathrm{HF}}^l \cdot \omega_k + \beta_{\mathrm{HF}}^l, \tag{4}$$

$$B_{k,\mathrm{HF}}^u = \alpha_{\mathrm{HF}}^u \cdot \omega_k + \beta_{\mathrm{HF}}^u. \tag{5}$$

Afterwards, the adaptive VLF⋆ band (0.01-0.05 Hz), LF⋆ band (from 0.05 Hz to B_{HF}^l), and HF⋆ band (from B_{HF}^l to B_{HF}^u) for the new subject can be obtained, yielding adaptive respiratory spectral features.

D. Feature Separability

As mentioned, we used the Mahalanobis distance MD to evaluate the separability of each single feature. Given a feature f, the MD value between the two classes *sleep* and *wake* for this feature is expressed as

$$\mathrm{MD}^{\mathrm{f}} = \frac{|\mu_{sleep}^{\mathrm{f}} - \mu_{wake}^{\mathrm{f}}|}{\sigma^{\mathrm{f}}}, \tag{6}$$

where μ^{f}_{sleep} and μ^{f}_{wake} are the population means of sleep and wake epochs, respectively; and σ^{f} is the standard deviation of the feature. With a single dimension, MD is also called the absolute standardized mean difference. A larger MD value corresponds to a higher separability.

E. Sleep/wake Detection

In this work, an LD classifier was used to perform sleep/wake detection, where more details of the LD classifier can be found elsewhere [5], [16]. Regarding the prior probabilities of the classifier, we observed that the probabilities of *sleep* and *wake* classes varied over time throughout the entire night. Thus, in order to exploit this variation that presented overnight, we computed the time-varying prior probability of each class for each epoch by computing the relative frequency of that specific epoch index annotated as each class [5], [14].

To assess the classification performance, we first considered overall accuracy (proportion of correctly identified epochs to the total number of epochs), specificity (proportion of correctly identified actual sleep epochs) and sensitivity (proportion of correctly identified actual wake epochs). In addition to these conventional criteria, we used a chance-adjusted metric the Cohen's Kappa coefficient [17] to evaluate the classification performance for the problem of "imbalanced class distribution" appearing in our data set, where the sleep epochs accounted for more than 90% during an entire-night recording. This metric allows for a better understanding of the general performance of a classifier in correctly identifying both classes.

A leave-one-subject-out cross-validation (LOOCV) procedure was used to perform the experiments for the purpose of evaluating the classifier. During each iteration of the LOOCV, 14 subjects were used to train the classifier and the remaining one was used for testing. Note that the optimal boundaries were estimated based on the training data and were then used to extract the features for the classification of the test data. This served to avoid biasing the classifier.

III. Results and Discussion

Figure 2 illustrates an example of estimating the optimal boundaries using the linear regression models. The mean regression coefficients over all subjects are shown in Table I. It indicates that the optimal boundaries can be linearly estimated with the mean respiratory frequency ω. We also found a significant Spearman's rank correlation r between B^l_{HF} and ω (r = 0.890, $p < 0.0001$) and between B^u_{HF} and ω (r = 0.893, $p < 0.0001$). It is noted that the adaptive boundaries were estimated only based on 14 data points (14 subjects) for each subject. This might not be sufficient to obtain an adequate linear model, therefore leading to errors in estimating the optimal boundaries. Thus, our method must be validated using a larger-sized data set in future work.

The average separabilities (over the 15 subjects) of the respiratory spectral features using the fixed boundaries (without adaptation) and using the adaptive boundaries (with adaptation) are compared in Table II. It shows that adapting

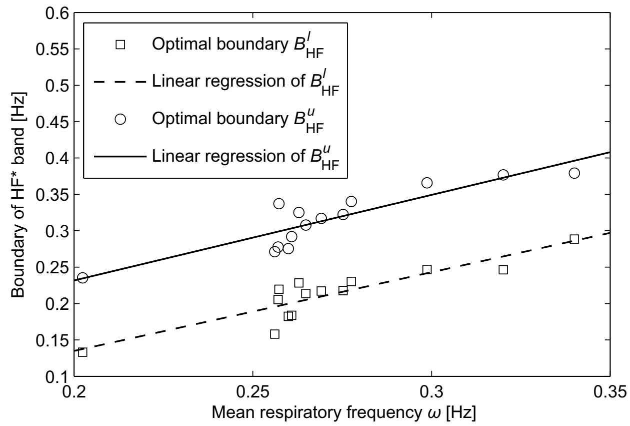

Fig. 2. An example of using the LSE linear regression method to estimate the adaptive boundaries for one subject based on the data from the other 14 subjects, where the R^2 values for fitting $B^l_{\mathrm{HF}} \sim \omega$ and $B^u_{\mathrm{HF}} \sim \omega$ curves are 0.803 and 0.798, respectively.

TABLE I

Average Estimates of Regression Coefficients using LOOCV

Boundary	α [Hz/Hz]	β [Hz]	R^2
B^l_{HF}	1.07 ± 0.09	-0.13 ± 0.02	0.74 ± 0.04
B^u_{HF}	1.49 ± 0.19	-0.08 ± 0.04	0.71 ± 0.04

the spectral boundaries of the respiratory spectral features using the proposed method resulted in significant increase in their separability, tested with a two-tailed paired Wilcoxon signed-rank test. For comparison, the table also presents the Mahalanobis distance of an actigraphy feature activity count (ACnt), the output of Actiwatch. It can be seen that, although the boundary adaptation can help increasing the separability of the respiratory spectral features, they are still relatively worse than ACnt that quantified body movements usually during wake state [18]. In fact, difficulty has been found in distinguishing between 'quiet' wake and REM sleep using respiratory activity [14].

The average classification results obtained using the fixed and the adaptive boundaries/bands are compared in Table III. Note that the classifier threshold was chosen to maximize Kappa coefficient. It is noticed that the use of adaptive spectral boundaries yielded significantly improved classification performances when using the respiratory spectral

TABLE II

Comparison of Average Feature Separability

Respiratory spectral feature	Feature separability (MD)	
	Fixed boundaries	Adaptive boundaries
VLF	1.23 ± 0.52	1.31 ± 0.49
LF	1.27 ± 0.60	1.42 ± 0.44
HF	1.47 ± 0.90	1.63 ± 0.79
LF/HF	1.33 ± 0.62	1.50 ± 0.49
ACnt	2.27 ± 0.99	

Note: The feature separability MD using adaptive boundaries was significantly higher than those using fixed boundaries for all the features (Wildoxon test, $p < 0.001$).

TABLE III

PERFORMANCE COMPARISON OF SLEEP/WAKE DETECTION WITH DIFFERENT FEATURE SETS USING LOOCV

Band	Sens. [%]	Spec. [%]	Acc. [%]	Kappa [%]
Actigraphy feature (ACnt)				
–	61.8 ± 16.1	97.3 ± 2.0	94.2 ± 1.7	0.57 ± 0.07
Respiratory spectral features (VLF, LF, HF, and LF/HF)				
Fixed	62.3 ± 21.5	88.2 ± 15.3	86.7 ± 12.9	0.42 ± 0.15
Adapt	60.1 ± 23.0	90.8 ± 12.2	89.0 ± 10.4*	0.45 ± 0.15*
All features (ACnt, VLF, LF, HF, and LF/HF)				
Fixed	68.4 ± 17.3	94.5 ± 9.0	92.8 ± 8.9	0.60 ± 0.14
Adapt	69.8 ± 17.1	94.3 ± 10.3	92.6 ± 9.2	0.62 ± 0.16

Note: Classifier operating thresholds were chosen to maximize Kappa coefficient. Significance of difference was examined between the results (accuracy and Kappa) obtained with the fixed and adaptive boundaries via a two-tailed Wilcoxon signed-rank test at $^*p < 0.05$.

TABLE IV

COMPARISON OF AVERAGE CORRELATION BETWEEN RESPIRATORY SPECTRAL FEATURES AND ACTIGRAPHY OVER SUBJECTS

Respiratory spectral feature	Spearman's correlation coefficient	
	Fixed boundaries	Adaptive boundaries
VLF	0.30 ± 0.02	0.30 ± 0.02
LF	0.27 ± 0.01	0.30 ± 0.06
HF	0.43 ± 0.04	0.48 ± 0.04
LF/HF	0.30 ± 0.02	0.34 ± 0.03

features alone (with a Kappa coefficient of 0.45 versus 0.42 and an accuracy of 89.0% versus 86.7%). As we expected, the results obtained using the respiratory spectral features (regardless of whether adapting the boundaries or not) are worse than those using the actigraphy feature ACnt. The classification results after combining the respiratory spectral features and ACnt are also presented in the table, where no significant difference was found between the results with and without boundaries adaptation. This might be because these features after boundary adaptation have higher (Spearman's rank) correlations with the actigraphy feature (see Table IV). It also indicates that adapting these features can help improving the detection of body movements conveyed by the respiratory effort signals to some extent. As shown in Figure 1, the spectral distribution during wake seems wider than that during sleep. This could either be a result of body movements or irregular breathing.

Since this study aimed at examining whether our boundary adaptation method can help improve the classification results, we only used the respiratory spectral features. In addition to them, many other existing respiratory (non-spectral) features have been used for sleep stage classification such as time domain features [5], amplitude features [6], and non-linear features based on self-similarity and sample entropy [7], [14]. Including those features for sleep/wake detection merits further investigation.

IV. CONCLUSION

To improve the performance of sleep/wake detection by reducing the between-subject variation conveyed in respiration, we achieved a boundary adaptation method to adapt the respiratory spectral features instead of using the traditional (fixed) boundaries. The adaptive boundaries were visually determined for each training subject so that they can maximize the separation of sleep and wake states. For each test subject, they were estimated using linear regression models based on those in training subjects. The boundary adaptation improved the discriminative power of the respiratory spectral features and therefore improved the sleep/wake detection. With an LD classifier tested with LOOCV, we obtained a significantly increased Kappa coefficient of 0.45 at an accuracy of 89.0% with the use of the respiratory spectral features alone. Moreover, we also obtained a slightly increased Kappa of 0.62 when combining them with actigraphy.

REFERENCES

[1] A. Rechtschaffen and A. Kales, A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington, DC: National Institutes of Health, 1968.

[2] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan, The AASM manual for the scoring of sleep and associated events: rules, terminology & technical specifications. Westchester, IL: The American Academy of Sleep Medicine, 2007.

[3] N. J. Douglas, D. P. White, C. K. Pickett, J. V. Weil, and C. W. Zwillich, Respiration during sleep in normal man, Thorax, vol. 37, no. 11, pp. 840–844, 1982.

[4] T. Penzel *et al.*, Cardiovascular and respiratory dynamics during normal and pathological sleep, Chaos, vol. 17, no. 1, 015116, 2007.

[5] S. J. Redmond and C. Heneghan, Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea, IEEE Trans. Biomed. Eng., vol. 53, no. 3, pp. 485–496, 2006.

[6] X. Long, J. Foussier, P. Fonseca, R. Haakma, and R. M. Aarts, Analyzing respiratory effort amplitude for automated sleep stage classification, Biomed. Signal Process. Control, vol. 14, pp. 197–205, 2014.

[7] X. Long, J. Yang, T. Weysen, R. Haakma, J. Foussier, P. Fonseca, and R. M. Aarts, Measuring dissimilarity between respiratory effort signals based on uniform scaling for sleep staging, Physiol. Meas., vol. 35, no. 12, pp. 2529–2542, 2014.

[8] L. Samy, M. -C. Huang, J. J. Liu, W. Xu, and M. Sarrafzadeh, Unobtrusive sleep stage identification using a pressure-sensitive bed sheet, IEEE Sens. J., vol. 14, no. 7, pp. 2092–2101, 2014.

[9] M. Zakrzewski, H. Raittinen, and J. Vanhala, Comparison of center estimation algorithms for heart and respiration monitoring with microwave Doppler radar, IEEE Sens. J., vol. 12, pp. 627–634, 2012.

[10] A. Heinrich, F. van Heesch, B. Puvvula, and M. Rocque, Video based actigraphy and breathing monitoring from the bedside table of shared beds, J. Ambient Intell. Human Comput., vol. 6, no. 1, pp. 107–120, 2015.

[11] S. Devot, D. Dratwa, and E. Naujokat, Sleep/wake detection based on cardiorespiratory signals and actigraphy, in Conf. Proc. IEEE Eng. Med. Biol. Soc., Buenos Aires, Argentina, 2010, pp. 5089–5092.

[12] Rostig *et al.*, Nonrandom variability of respiration during sleep in healthy humans, Sleep, vol. 28, no. 4, pp. 411–417, 2005.

[13] P. C. Mahalanobis, On the generalized distance in statistics, Proc. Nat. Inst. Sci. India, vol. 2, pp. 49–55, 1936.

[14] X. Long, P. Fonseca, J. Foussier, R. Haakma, and R. M. Aarts, Sleep and wake classification with actigraphy and respiratory effort using dynamic warping, IEEE J. Biomed. Health Inform., vol. 18, no. 4, pp. 1272–1284, 2014.

[15] J. H. Pollard, A Handbook of Numerical and Statistical Techniques. New York: Cambridge University Press, 1977.

[16] X. Long, P. Fonseca, R. Haakma, R. M. Aarts, and J. Foussier, Spectral boundary adaptation on heart rate variability for sleep and wake classification, Int. J. Artif. Intell. Tools, vol. 23, no. 3, 1460002, 2014.

[17] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas., vol. 20, pp. 37–46, 1960.

[18] A. Sadeh and C. Acebo, The role of actigraphy in sleep medicine, Sleep Med. Rev., vol. 6, no. 2, pp. 113–124, 2002.