Biomedical Signal Processing and Control 14 (2014) 197–205

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

# Analyzing respiratory effort amplitude for automated sleep stage classification

Xi Long[a,b,*], Jérôme Foussier[c], Pedro Fonseca[a,b], Reinder Haakma[b], Ronald M. Aarts[a,b]

[a] Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands
[b] Personal Health Group, Philips Group Innovation Research, The Netherlands
[c] Philips Chair for Medical Information Technology (MedIT), RWTH Aachen University, Germany

ARTICLE INFO

*Article history:*
Received 6 May 2014
Received in revised form 10 July 2014
Accepted 1 August 2014
Available online 27 August 2014

*Keywords:*
Respiratory effort amplitude
Signal calibration
Feature extraction
Sleep stage classification

ABSTRACT

Respiratory effort has been widely used for objective analysis of human sleep during bedtime. Several features extracted from respiratory effort signal have succeeded in automated sleep stage classification throughout the night such as variability of respiratory frequency, spectral powers in different frequency bands, respiratory regularity and self-similarity. In regard to the respiratory amplitude, it has been found that the respiratory depth is more irregular and the tidal volume is smaller during rapid-eye-movement (REM) sleep than during non-REM (NREM) sleep. However, these physiological properties have not been explicitly elaborated for sleep stage classification. By analyzing the respiratory effort amplitude, we propose a set of 12 novel features that should reflect respiratory depth and volume, respectively. They are expected to help classify sleep stages. Experiments were conducted with a data set of 48 sleepers using a linear discriminant (LD) classifier and classification performance was evaluated by overall accuracy and Cohen's Kappa coefficient of agreement. Cross validations (10-fold) show that adding the new features into the existing feature set achieved significantly improved results in classifying wake, REM sleep, light sleep and deep sleep (Kappa of 0.38 and accuracy of 63.8%) and in classifying wake, REM sleep and NREM sleep (Kappa of 0.45 and accuracy of 76.2%). In particular, the incorporation of these new features can help improve deep sleep detection to more extent (with a Kappa coefficient increasing from 0.33 to 0.43). We also revealed that calibrating the respiratory effort signals by means of body movements and performing subject-specific feature normalization can ultimately yield enhanced classification performance.



## 1. Introduction

According to the rules presented by Rechtschaffen and Kales (the R&K rules) [1], human sleep is comprised of wake, rapid-eye-movement (REM) sleep and four non-REM (NREM) sleep stages S1–S4. S1 and S2 are usually grouped as "light sleep" and S3 and S4 correspond to slow-wave sleep (SWS) or "deep sleep" [2]. The gold standard for nocturnal sleep assessment is overnight polysomnography (PSG) which is typically collected in a sleep laboratory. With PSG, sleep stage is manually scored on each 30-s epoch throughout the night by trained sleep experts, forming a sleep hypnogram [1]. PSG recordings usually contain multiple bio-signals such as electroencephalography (EEG), electrocardiography (ECG), electrooculography (EOG), electromyography (EMG), respiratory effort, and blood oxygen saturation.

Respiratory information has been widely used for objectively assessing human nocturnal sleep [3–5]. Detecting sleep stages overnight is beneficial to the interpretation of sleep architecture or monitoring of sleep-related disorders [6,7]. Cardiorespiratory-based automated sleep stage classification has been increasingly studied in recent years [8–12]. Some of those studies only made use of respiratory activity because, when comparing with it cardiac activity is relatively more difficult to be captured reliably in an unobtrusive manner [10,11]. For respiratory activity, in comparison with the breathing ventilation acquired with traditional devices such as nasal prongs or face mask [13], respiratory effort can be obtained in an easier and more noninvasive or unobtrusive way, e.g., using a respiratory inductance plethysmography (RIP) sensor [14], an infrared (IR) camera [15], or a pressure sensitive bed-sheet [16].

Several parameters have been derived from respiratory effort signals for sleep analysis including respiratory frequency, powers of different respiratory spectral bands [8], respiratory self-similarity

* Corresponding author at: Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Den Dolech 2, 5612 AE Eindhoven, The Netherlands. Tel.: +31 6 8647 1335.
*E-mail addresses:* xi.long@philips.com, x.long@tue.nl, xi.long.ee@gmail.com (X. Long).

http://dx.doi.org/10.1016/j.bspc.2014.08.001
1746-8094/

[11], regularity [17] etc. These parameters are usually called "features" in the tasks of epoch-by-epoch sleep stage classification. In addition, it has been reported that the respiratory amplitude (e.g., depth and volume) differs between sleep stages [4]. For instance, the "respiratory depth" is more regular and the tidal volume, minute ventilation, and inspiratory flow rate are significantly lower during REM sleep than during NREM sleep (particularly during deep sleep) [18,19]. To the authors' knowledge, these characteristics that express different physiological properties across sleep stages have not been explicitly elaborated and quantified for applications of sleep stage classification. We therefore exploit these characteristics by analyzing respiratory effort signal envelope and area. Features quantifying these characteristics are motivated to be designed which are expected to in turn help separate different sleep stages.

It is assumed that the information about respiratory depth or volume is obtainable from the respiratory effort signal. For instance, the signal (upper and lower) envelopes and area should correspond to respiratory depth and volume, respectively. In fact, respiratory effort has often been used as a surrogate of tidal volume since it is obtained by measuring motions of rib cage or abdominal with, e.g., RIP [14]. However, Whyte et al. [20] argued that this assumption does not always hold, particularly when a sleeper changes his/her posture along with body movements during sleep. This is because the respiratory effort amplitude might be affected by body movements as the sensor position may shift and/or the sensor may be stretched. This will cause an uneven comparison of the signal amplitude before and after body movements, yielding errors when computing the feature values. In order to provide a more accurate estimate of respiratory depth and volume from respiratory effort signal, we must calibrate the signal by means of body movements. They can be quantified by analyzing the artifacts of respiratory effort signal (often inline with body movements) using a dynamic time warping (DTW)-based method [11]. DTW is a signal-matching algorithm that quantifies an optimal non-linear alignment between two time series allowing scaling and offset [21]. Our previous work [11] has proposed a DTW measure to effectively capture body motion artifacts by measuring self-similarity of respiratory effort. This measure has been successfully used as a feature for classifying sleep and wake states in that work. Therefore, we simply adopted this measure to detect motion artifacts modulated by body movements in respiratory effort signals. Using the DTW-based method enables the exclusion of an additional sensor modality (e.g., actigraphy) specifically used for detecting body movements.

The address of this paper is exclusively on investigating a set of novel features that can characterize respiratory amplitude in different aspects with the ultimate goal of improving sleep stage classification performance. Previous studies have shown that linear discriminant (LD) is an appropriate algorithm in sleep stage classification [6,8,22]. Likewise, we simply adopted an LD classifier. Preliminary results of this work in classifying REM and NREM sleep have been previously published [23].

## 2. Materials and methods

### 2.1. Subjects and data

Data of 48 healthy subjects (21 males and 27 females) in the SIESTA project (supported by European Commission) [24] were included in our data set. The subjects had a Pittsburgh Sleep Quality Index (PSQI) of no more than 5 and met several criteria (no shift work, no depressive symptoms, usual bedtime before midnight, etc.). All the subjects signed an informed consent form prior to the study, documented their sleep habits over 14 nights, and underwent overnight PSG study for two consecutive nights (on day 7 and day 8) in sleep laboratories. The PSG recordings collected on day 7 were used for analyses, from which the respiratory effort signals (sampling rate of 10 Hz) were recorded with thoracic inductance plethysmography.

**Table 1**
Summary of subject demographics and some sleep statistics ($N = 48$).

| Parameter | Mean ± SD | Range |
|---|---|---|
| Sex | 21 males and 27 females | |
| Age (years) | 41.3 ± 16.1 | 20–83 |
| BMI[a] ($kg\,m^{-2}$) | 23.6 ± 2.9 | 19.1–31.3 |
| TRT[b] (h) | 7.8 ± 0.4 | 6.6–8.6 |
| Wake, W (%) | 12.9 ± 6.1 | 1.2–24.5 |
| REM sleep, R (%) | 19.0 ± 3.3 | 15.3–26.5 |
| NREM sleep, N (%) | 68.1 ± 4.9 | 56.1–76.3 |
| Light sleep, L (%) | 53.6 ± 5.5 | 42.7–66.7 |
| Deep sleep, D (%) | 14.5 ± 4.8 | 5.3–28.5 |

[a] Body mass index.
[b] Total recording time.

Sleep stages were manually scored on 30-s epochs as wake, REM sleep, or one of the NREM sleep stages by sleep clinicians based on the R&K rules. For sleep stage classification epochs were labeled as four classes W (wake), R (REM sleep), L (light sleep), and D (deep sleep), or three classes W, R, and N (NREM sleep).

From the data used in this study the subject demographics and some sleep statistics [mean ± standard deviation (SD) and range] are summarized in Table 1.

### 2.2. Signal preprocessing

The raw respiratory effort signals of all subjects were preprocessed before feature extraction. They were filtered with a 10th order Butterworth low-pass filter with a cut-off frequency of 0.6 Hz for the purpose of eliminating high frequency noise. Afterwards the baseline was removed by subtracting the median peak-to-trough amplitude. To locate the peaks and troughs, we identified the turning points simply based on sign change of signal slope and then corrected the falsely detected 'dubious' peaks and troughs (1) with too short intervals between peak and trough pairs where the sum of two successive intervals is less than the median of all intervals over the entire recording and (2) with two small amplitudes where the peak-to-trough difference is smaller than 15% of the median of the entire respiratory effort signal. These methods were validated by comparing automatically detected results with manually annotated peaks and troughs and an accuracy of ~98% was achieved.

### 2.3. Existing respiratory features

A pool of 14 existing features extracted from the respiratory effort signal has been used in previous studies for sleep stage classification. In the time domain, the mean and SD of breath lengths ($L_m$ and $L_{sd}$) and the mean and SD of breath-by-breath correlations ($C_m$ and $C_{sd}$) were calculated [6]. In the frequency domain, we extracted features based on the respiratory effort spectrum for each epoch where the spectrum was estimated using a short time Fourier transform (STFT) with a Hanning window. From the spectrum the dominant frequency ($F_r$) in the range of 0.05–0.5 Hz (estimated as the respiratory frequency) and the logarithm of its power ($F_p$) were obtained [6]. We also took the logarithm of the spectral power in the very low frequency band between 0.01 and 0.05 Hz (VLF), low frequency band between 0.05 and 0.15 Hz (LF), and high frequency band from 0.15 to 0.5 Hz (HF) and the ratio between LF and HF spectral powers (LF/HF) [6,8]. Furthermore the standard deviation of respiratory frequency over 5 epochs ($F_{sd}$) was computed [8]. Non-linear features consist of self-similarity measured