Measuring dissimilarity between respiratory effort signals based on uniform scaling for sleep staging

2014 Physiol. Meas. 35 2529

(http://iopscience.iop.org/0967-3334/35/12/2529)

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **35** (2014) 2529–2542 doi:10.1088/0967-3334/35/12/2529

Measuring dissimilarity between respiratory effort signals based on uniform scaling for sleep staging

Xi Long[1,2], Jie Yang[3], Tim Weysen[2], Reinder Haakma[2], Jérôme Foussier[4], Pedro Fonseca[1,2] and Ronald M Aarts[1,2]

[1] Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands
[2] Philips Group Innovation Research, 5656 AE Eindhoven, The Netherlands
[3] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands
[4] Chair for Medical Information Technology, RWTH Aachen University, 52074 Aachen, Germany

E-mail: xi.long@philips.com and xi.long.ee@gmail.com

Received 26 March 2014, revised 29 August 2014
Accepted for publication 17 September 2014
Published 19 November 2014

Abstract

Polysomnography (PSG) has been extensively studied for sleep staging, where sleep stages are usually classified as wake, rapid-eye-movement (REM) sleep, or non-REM (NREM) sleep (including light and deep sleep). Respiratory information has been proven to correlate with autonomic nervous activity that is related to sleep stages. For example, it is known that the breathing rate and amplitude during NREM sleep, in particular during deep sleep, are steadier and more regular compared to periods of wakefulness that can be influenced by body movements, conscious control, or other external factors. However, the respiratory morphology has not been well investigated across sleep stages. We thus explore the dissimilarity of respiratory effort with respect to its signal waveform or morphology. The dissimilarity measure is computed between two respiratory effort signal segments with the same number of consecutive breaths using a uniform scaling distance. To capture the property of signal morphological dissimilarity, we propose a novel window-based feature in a framework of sleep staging. Experiments were conducted with a data set of 48 healthy subjects using a linear discriminant classifier and a ten-fold cross validation. It is revealed that this feature can help discriminate between sleep stages, but with an exception of separating wake and REM sleep. When combining the new feature with 26 existing respiratory features, we achieved a Cohen's Kappa coefficient of 0.48 for 3-stage classification (wake, REM sleep and NREM sleep) and of 0.41 for 4-stage classification (wake, REM

sleep, light sleep and deep sleep), which outperform the results obtained without using this new feature.

Keywords: sleep staging, respiratory effort, dissimilarity measure, feature extraction, uniform scaling

(Some figures may appear in colour only in the online journal)

1. Introduction

Previous studies have shown that characteristics of human respiratory activity are associated with sleep stages throughout the entire night (Douglas *et al* 1982, Somers *et al* 1993). Respiratory effort has been increasingly used for objective sleep analysis (Roebuck *et al* 2014) and sleep staging (Redmond *et al* 2007, Chung *et al* 2009) in contrast to traditional polysomnography (PSG) which is considered the 'gold standard' in sleep studies. This is because respiratory activity is able to be acquired in an easy and unobtrusive manner using, for example, bed sensors (Watanabe *et al* 2005, Kortelainen *et al* 2010), Doppler radar (Matthews *et al* 2000), photoplethysmography (Lázaro *et al* 2013), or a watch-based device (Herscovici *et al* 2007). Sleep consists of wake, rapid-eye-movement (REM) sleep and four non-REM (NREM) sleep stages S1–S4 according to the R&K rules (Rechtschaffen and Kales 1968). In regard to S3 and S4, the American Academy of Sleep Medicine (AASM) guidelines (Iber *et al* 2007) and their updated rules (Berry *et al* 2012) suggest merging them into a single 'deep sleep' or slow wave sleep stage. S1 and S2 often correspond to 'light sleep' (Silber *et al* 2007, Bresler *et al* 2008). With PSG, sleep stages are manually scored by sleep technicians on 30 s epochs based on multiple electrophysiological signals including electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG). The manually scored sleep stages can be visualized in a hypnogram.

It has been reported in earlier studies that some characteristics of respiration differ across sleep stages such as respiratory frequency (Douglas *et al* 1982), respiratory variability (Rostig *et al* 2005), different frequency components of respiratory spectrum (Redmond *et al* 2007), etc. However, the dissimilarity of respiratory effort in terms of signal waveform or morphology for different sleep stages has not been well explored. In fact, the respiratory pattern (e.g. amplitude and frequency) has been shown to be more stable and regular during NREM sleep (in particular during deep sleep) than during wake and REM sleep (Cherniack 1981, Heinzer and Sériès 2011). The irregularity of breathing is usually caused by body movements, alternation of ventilation control, or behavioral factors when awake (Phillipson 1978) and it is related to paralysis of voluntary musculature (muscle atonia) during REM sleep (Polkey *et al* 1995). In this matter, we may then anticipate that if a sleep stage has a higher regularity in breathing, the respiratory effort in this stage would have lower dissimilarity in between. On the other hand, the respiratory dynamics have been found to associate with physiologic states such as sleep stages which distinctly correspond to autonomic regulatory mechanisms (Trinder *et al* 2001, Penzel *et al* 2007, Schumann *et al* 2010). We therefore hypothesise that (1) the respiratory effort is characterized by signal morphology and (2) the dissimilarity between two respiratory effort periods is influenced by their corresponding sleep stages. Research has been focusing on investigating respiration changes during sleep (Kantelhardt *et al* 2003, Rostig *et al* 2005). For instance, some researchers analyzed non-random variability of respiration (e.g. breath-by-breath intervals) on short- and long-term scales (Rostig *et al* 2005), whereas with a much less focus on

comparing respiratory patterns of multiple breaths. Although some parameters including breathing rate, inspiratory/expiratory volumes and minute volume were investigated, the respiratory morphology was less researched. Many methods have been utilized to compare two time series such as cross-correlation, detrended fluctuation analysis and cross-approximate entropy, however, they can be limited by several factors including the non-stationary trend of data, insufficient number of data points for, e.g. polynomial fitting, low relative consistency and/or unequal length between time series (Richman and Moorman 2000, Bashan *et al* 2008, Horvatic *et al* 2011). The idea here is to use a Euclidean-based distance as a dissimilarity metric between two respiratory effort signal segments from a subject. When computing the distance, each signal segment is selected inside its corresponding 30 s epoch to have a certain number of consecutive breaths, served to provide an even comparison on their signal morphology. These signal segments are usually less than 30 s. It is inevitable that the length (i.e. number of data points) of any two signal segments differs so that they are necessarily required to be scaled at an equal length in order to perform an Euclidian (sequential) mapping. To resolve this problem, we propose to use a uniform scaling method (Yankov *et al* 2007) to re-scale the two signal segments by searching for the minimal Euclidean distance between them. In other words, they are uniformly 'stretched' to allow for a reduction on the effects of variant breathing frequency to a certain degree, resulting in focusing more on signal morphology.

As for automatic sleep staging, it is particularly interesting to know if different sleep stages can be distinguished by means of respiratory effort data when the PSG-based hypnogram is absent. This would benefit the applications of home-based sleep staging or sleep stage classification which has been attracting increasing attention in recent years (Redmond and Heneghan 2006, Devot *et al* 2010, Long *et al* 2014a, Samy *et al* 2014). Information regarding sleep stages is usually extracted as epoch-based 'features' used to perform epoch-by-epoch classification. For this purpose, we propose a new feature to describe the dissimilarity of respiratory effort morphology between different epochs from the same recording. Of this feature, discriminative power in classifying sleep stages will be evaluated and it is expected to help improve sleep staging performance.

2. Materials and methods

2.1. Subjects and protocol

Forty eight healthy subjects (21 men and 27 women; mean age 41.3 years ranging from 20 to 83, standard deviation (SD) 16.1; mean body mass index 23.6 $kg{\cdot}m^{-2}$ ranging from 19.1 to 31.3, SD 2.9) in the SIESTA project (Klösch *et al* 2001) are considered. The project was supported by the European Commission and the subjects were monitored in seven different sleep laboratories located in five European countries over a period of three years from 1997 to 2000. The subjects had a Pittsburgh Sleep Quality Index (Buysse *et al* 1989) of less than 6 and fulfilled several criteria (e.g. no depressive symptoms, no reported medical, neurological, mental or cardiovascular disorders, no history of drug abuse or habituation, no psychoactive medication, no shift work and usually bedtime before midnight). According to the study protocol of the SIESTA project, all subjects provided an informed consent, documented their sleep habits over 14 nights and spent two consecutive nights (on days 7 and 8) in the sleep laboratory (Anderer *et al* 2005). More details regarding the subject information and the study protocol can be found online (www.ofai.at/siesta). In this study, we only include single-night PSG recordings (on day 7) for analysis.

Table 1. Sleep data from 48 healthy subjects, where mean ± SD and range are given.

Parameter	Mean ± SD	Range
Total recording time (hours)	7.8 ± 0.4	6.6–8.6
Total number of epochs (#)	938.3 ± 44.5	796–1026
Wake (%)	12.9 ± 6.1	1.2–24.5
REM sleep (%)	19.0 ± 3.3	15.3–26.5
NREM sleep (%)	68.1 ± 4.9	56.1–76.3
Light sleep (%)	53.6 ± 5.5	42.7–66.7
Deep sleep (%)	14.5 ± 4.8	5.3–28.5

2.2. Polysomnographic measurements

Full PSG data, including multiple EEG-, EOG- and EMG-channels, electrocardiography (ECG), respiratory effort, oxygen saturation, snoring, etc were recorded for each subject and the sleep stages were visually scored by professional sleep technicians as wake, REM and S1–S4 on 30 s epochs according to the R&K rules. Thoracic breathing movements were measured by respiratory inductance plethysmography (RIP) in the form of respiratory effort signals at a sampling rate of 10 Hz. For the problem of sleep staging, we consider deep sleep (merged S3 and S4) as a single stage as suggested by the AASM guidelines. In the mean time, S1 and S2 are merged as single light sleep.

Referring to the statistics of normal sleepers across the human lifespan reported previously (Ohayon *et al* 2004), the selection of overnight recordings from a larger data set met several criteria including the sleep efficiency of $\geqslant$75%, REM sleep of $\geqslant$15% and deep sleep of $\geqslant$5%. The sleep data is summarized in table 1, in which mean and SD over subjects and range are presented.

2.3. Signal processing

The raw respiratory effort signals are first low-pass filtered (10th order Butterworth filter with a cut-off frequency of 0.6 Hz) in order to eliminate high-frequency noise. Then the baseline is removed by subtracting the median peak-to-trough amplitude estimated over the entire recording, which serves to compute the respiratory volume-based features. These features will be described further in section 2.7. The localization of respiratory peaks/troughs is achieved by detecting the signal turning points based on sign changes of the signal slopes. Afterwards, we remove the falsely detected peaks/troughs (1) with too short peak-to-trough or trough-to-peak intervals (where the sum of two successive intervals is less than the median of all intervals over the entire recording) and (2) with too small amplitudes (where the peak-to-trough difference is smaller than 0.15 times the median of the entire respiratory signal). These methods were validated by comparing the automatically detected results with manually annotated peaks and troughs and an accuracy of ~98% was achieved.

2.4. Dissimilarity measure with uniform scaling

Given an overnight respiratory effort recording with L epochs from a subject, the ith epoch is expressed as $U_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,n}\}(i = 1, 2, \ldots, L)$ with n data points (here $n = 300$ at the signal sampling rate of 10 Hz). As explained before, we only choose a signal segment with a certain number of consecutive breaths λ inside this epoch when computing the

dissimilarity score, thereby the chosen signal segment for this particular epoch U_i is expressed by $V_i = \{v_{i,1}, v_{i,2}, \ldots, v_{i,m_i}\}$ with m_i data points ($m_i \leqslant n$). The locations of $v_{i,1}$ and v_{i,m_i} are based on the detected respiratory peaks or troughs within this epoch so that the segment V_i contains several complete breaths, starting and ending at two different troughs. The signal segment length m_i is dependent of i because respiratory frequency usually varies between signal segments, even if they might have a same number of breaths. Besides, it also depends on the prescribed number of breaths λ.

Let us consider two epochs U_i and $U_j (i, j = 1, 2, \ldots, L$ and $i \neq j)$ with p_i and q_i consecutive breaths, respectively. To ensure an equal number of breaths that aims at evenly comparing their dissimilarity, we have $\lambda = \min\{p_i, q_i\}$. For the epoch with more breaths, only the λ breaths in the middle are selected, yielding a signal segment within this epoch. Then the two signal segments V_i and $V_j (i \neq j)$ with λ breaths each are normalized at zero mean and unit variance (Z-score normalization). However, the two signal segments may have unequal lengths, which is not applicable for computing the Euclidean distance between them. To tackle this, we utilize uniform scaling, a Euclidean-based minimization method. For V_i and V_j, assuming that $m_i \leqslant m_j$, a uniformly scaled series of V_i is expressed as $V_i^k = \{v_{i,1}^k, v_{i,2}^k, \ldots, v_{i,k}^k\}$ with length of $k (m_i \leqslant k \leqslant m_j)$, where $v_{i,x}^k = v_{i,\lceil x.m_i.k^{-1} \rceil}$ for $x = 1, 2, \ldots, m_i$. Hence, the dissimilarity score d_{score} between U_i and U_j is the uniform scaling distance d_{us} between V_i and V_j, which can be obtained by minimizing the Euclidean distance subject to $m_i \leqslant m_j$, such that

$$d_{\text{score}}(U_i, U_j) \equiv d_{\text{us}}(V_i, V_j) = \min_{m_i \leqslant k \leqslant m_j} \sqrt{\frac{1}{k} \sum_{x=1}^{k} (v_{i,x}^k - v_{j,x})^2}. \tag{1}$$

Since the k-space Euclidian distance metric is sensitive to series length k which usually encounters different values in equation (1), the distance should be normalized by k. Figure 1 depicts an example of computing the dissimilarity score d_{score} between two epochs. Note that d_{score} is computed within each recording (or subject for the single-night data) to avoid the effect of between-subject variability, often caused by the existence of physiological difference from subject to subject.

2.5. Windowed dissimilarity feature

It is of interest to extract a feature for each 30 s epoch to capture the dissimilarity property of respiratory effort morphology. This feature can in turn be used to separate different sleep stages. To do so, we compute the mean dissimilarity score between each epoch and the other epochs from the same recording within a window, named by windowed (self-) dissimilarity feature and denoted as D_{win} henceforth. We expect that this feature is not independent of sleep stage and thus it is informative for sleep staging. For the ith epoch U_i from a given subject, it is computed as

$$\text{D}_{\text{win}}(U_i) = \frac{\sum_{j=1}^{L} d_{\text{score}}(U_i, U_j)}{\min(w, i-1) + \min(w, L-i)}, \quad \text{for } |j - i| \leqslant w \text{ and } j \neq i, \tag{2}$$

in which L is the total number of epochs for this specific subject and $w = 1, 2, \ldots, L$ is the (single-side) size of the window centered at U_i. This means that D_{win} is a feature with a certain time (or window) scale. The window size w is determined by maximizing the feature discriminative power. Intuitively, the majority of the epochs contained within a small window should be in the same sleep stage as the given epoch. This can be examined by comparing

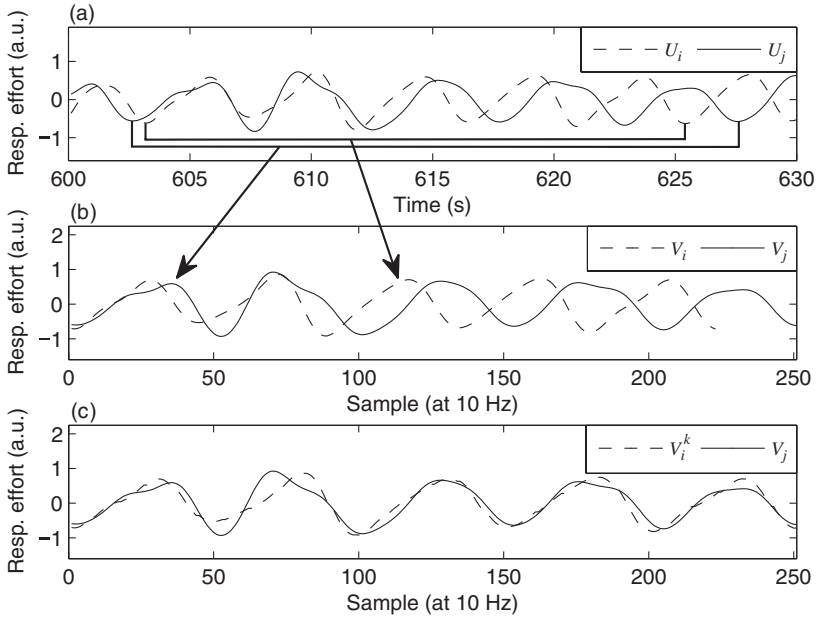

Figure 1. An example of computing the dissimilarity score of respiratory effort between two epochs: (*a*) original signals U_i and U_j at 10 Hz within 30 s epochs; (*b*) selected signal segments V_i and V_j with 5 consecutive breaths, where series lengths are unequal; (*c*) uniformly scaled series V_i^k and V_j, where k equals the length of V_j. Note that the signal segments in (*a*) and (*b*) are normalized to have zero mean and unit variance.

the percentage of occurrence for different sleep stages versus the time difference Δ between epochs. We also analyze the changes of d_{score} for 'self-comparisons' versus Δ, where d_{score} is computed between epochs with same sleep stage (i.e. wake–wake, REM–REM, light–light and deep–deep). To reduce noise in feature level caused by measurement errors or body motion artifacts, D_{win} is smoothed over the entire-night recording using a moving average method (with a 10 min span).

2.6. Feature analysis

For the windowed dissimilarity feature D_{win}, we first compare its mean value and SD over all subjects between sleep stages. In addition to that, we compute its discriminative power for sleep staging using One-Way analysis of variance (ANOVA) F-statistic. A higher discriminative power leads to a larger value of ANOVA F-statistic. The F-statistic of D_{win} is then compared with that of the existing features by ranking it among all the features. The distributions of D_{win} in different sleep stages are found to approximately follow a normal distribution using a Quantile–Quantile (QQ) plot method.

2.7. Sleep staging

As stated, the new feature D_{win} can be incorporated to perform automatic sleep staging when solely using respiratory effort data. A set of 26 existing respiratory features have been used to classify sleep stages in previous studies. They comprise features in both time and frequency domain (Redmond and Heneghan 2006), respiratory depth- and volume-based features (Long

Table 2. A list of respiratory features.

Feature Index	Description
1	Respiratory frequency estimated in the frequency domain
2	Spectral power of respiratory frequency
3	Spectral power in very low frequency (VLF) band (0.01–0.05 Hz)
4	Spectral power in low frequency (LF) band (0.05–0.15 Hz)
5	Spectral power in high frequency (HF) band (0.15–0.5 Hz)
6	Ratio of spectral power between LF and HF bands
7	Standard deviation of respiratory frequency over 150 s
8	Mean breath-by-breath correlation
9	Standard deviation of breath-by-breath correlation
10	Standard deviation of breath length
11	Respiratory frequency estimated in the time domain
12	Respiratory regularity measured by sample entropy
13	Respiratory similarity measured by dynamic time warping
14	Respiratory similarity measured by dynamic frequency warping
15	Standardized median of respiratory peaks
16	Standardized median of respiratory troughs
17	Respiratory peak regularity measured by sample entropy
18	Respiratory trough regularity measured by sample entropy
19	Median respiratory peak-to-trough difference
20	Median respiratory volume during breath cycles
21	Median respiratory volume during inhalations
22	Median respiratory volume during exhalations
23	Median respiratory flow rate during breath cycles
24	Median respiratory flow rate during inhalations
25	Median respiratory flow rate during exhalations
26	Ratio of inhalation and exhalation flow rate
27	Respiratory dissimilarity measured by uniform scaling (D_{win})

Note: the references for the existing features are 1–11 (Redmond and Heneghan 2006, Redmond *et al* 2007), 12 (Costa *et al* 2008), 13 and 14 (Long *et al* 2014a) and 15–26 (Long *et al* 2014b).

et al 2014b) and non-linear features based on sample entropy (Costa *et al* 2008) and dynamic warping (Long *et al* 2014a). Table 2 lists and describes all the respiratory features. To examine whether D_{win} can help achieve an enhanced classification performance, we compare the classification results with and without adding it to the existing feature set. Note that for the purpose of reducing between-subject variability in respiration, all the features are normalized (Z-score) for each overnight recording.

We simply adopt a linear discriminant (LD) classifier which has been widely used for the task of sleep staging (Redmond *et al* 2007, Devot *et al* 2010, Foussier *et al* 2013, Long *et al* 2014a). The data including 48 entire-night recordings is randomly divided to 10 data subsets where each fold consists of four or five recordings and then we execute the sleep staging iteratively using a ten-fold cross-validation (CV). During each iteration, the classifier is trained on nine folds and validated on the remaining one in order to minimize the classifier bias.

To evaluate the classifier, we use Cohen's Kappa coefficient κ (Cohen 1960) in addition to overall accuracy because it is more appropriate for analyzing unbalanced data (in our case light sleep accounts for 53.6% which is much larger than the other stages). To exploit the prior probabilities of different sleep stages in an LD classifier that may change over time, we compute a time-varying prior probability (TVPP) for each epoch by counting the relative

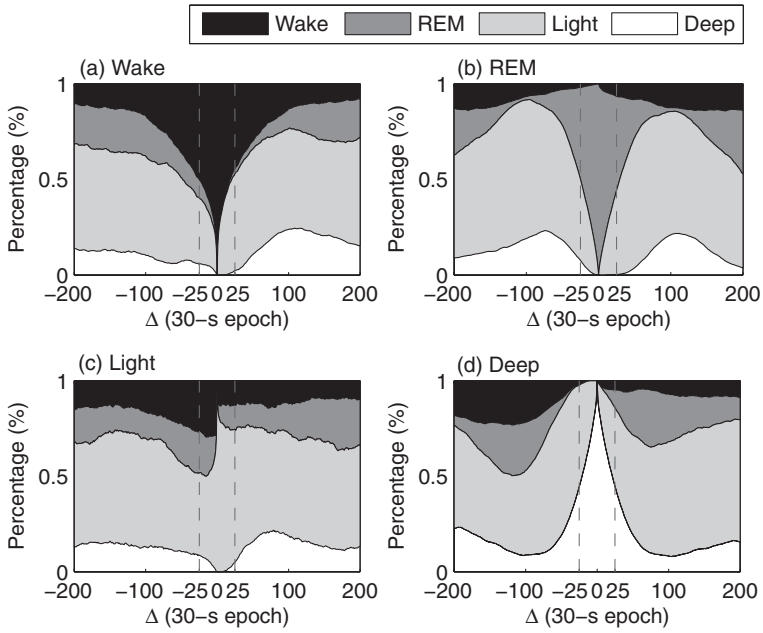

Figure 2. The probability of occurrence of different sleep stages versus time difference Δ for (*a*) wake, (*b*) REM, (*c*) light and (*d*) deep sleep epochs. It can be seen that, for all these stages, light sleep percentage (filled in light gray) is larger than any other stages when $|\Delta| > \sim$ 30 epochs. The boundary of the 25-epoch window for computing D_{win} is indicated (dashed line).

frequency of occurrence of each sleep stage at its corresponding time of the night based on the associated training data. More details about TVPP can be found elsewhere (Redmond *et al* 2007). Here we present results for two sleep staging schemes, including 4-stage classification (wake, REM sleep, light sleep and deep sleep) and 3-stage classification (wake, REM sleep and NREM sleep).

3. Results

The (single-side) window size w of 25 epochs was experimentally found to be an appropriate value when computing the new feature D_{win}, where its feature discriminative power in classifying wake, REM sleep, light sleep and deep sleep was maximized. Figure 2 compares the percentage of occurrence in different sleep stages changing over Δ. The figure indicates a presence of self-comparisons with a higher likelihood if $|\Delta|$ is smaller than a value (e.g. ∼30 epochs for wake, REM sleep and deep sleep). It also illustrates that the comparison between each sleep stage and light sleep dominates if $|\Delta|$ is larger than that value. These graphs imply that, for our choice of w = 25 epochs, the feature values of D_{win} depend more on the self-comparisons. As shown in figure 3, in regard to the self-comparisons, we observe that different sleep stages can be separated by the dissimilarity score within the 25-epoch window except for that between wake and REM sleep where overlaps occur.

Figure 4 compares the feature values of D_{win} in different sleep stages (mean ± SD and histogram), in which the separation can be observed between sleep stages, particularly between deep sleep and the other stages and between REM and NREM sleep. An example of an

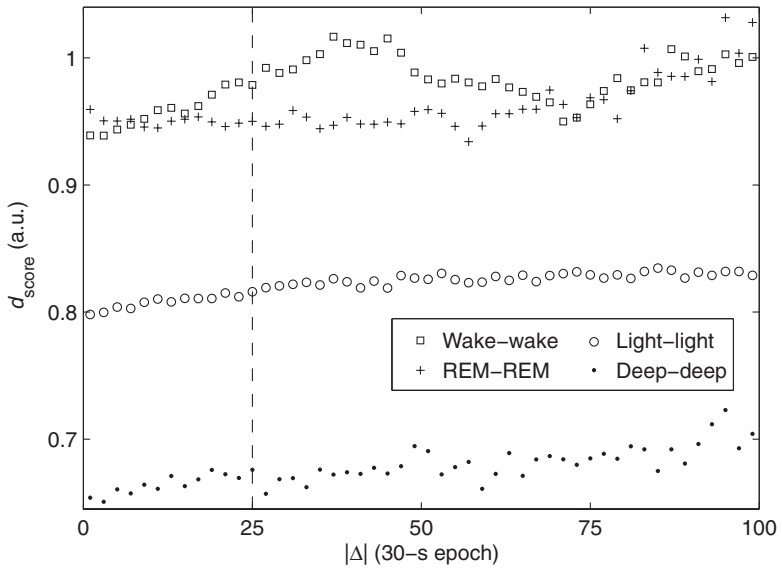

Figure 3. Mean dissimilarity score d_{score} versus absolute time difference |Δ| for self-comparisons wake–wake, REM–REM, light–light and deep–deep. The boundary of the 25-epoch window for computing D_{win} is indicated (dashed line).

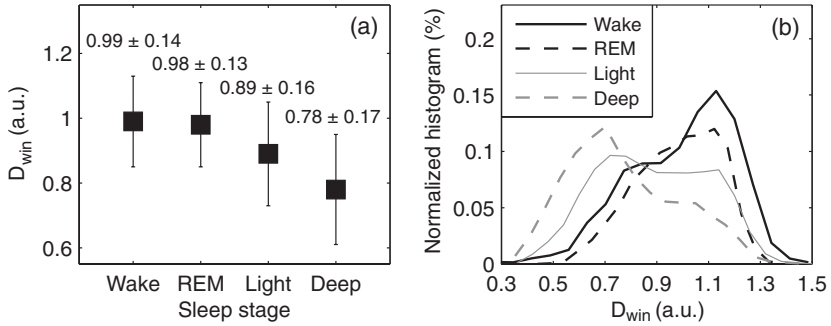

Figure 4. Comparison of the windowed dissimilarity feature D_{win} in different sleep stages: (*a*) mean ± SD and (*b*) normalized histogram (i.e. percentage, %).

overnight hypnogram and the corresponding D_{win} values from a 50 year-old female are illustrated in figure 5, where the correlation between them can be seen. Table 3 presents the discriminative powers (as measured by ANOVA F-statistic) of D_{win} in separating different sleep stages. For comparison, we also provide its ranking among all features as well as the top-10 ranked features (in a descending order in terms of F-statistic) in the table.

The respiratory effort-based sleep staging results using the feature set with and without D_{win} are compared in table 4, where the overall accuracy and the Cohen's Kappa coefficient are reported. It is noted that combining D_{win} with the existing features resulted in a significantly increased κ of 0.41 at an overall accuracy of 64.9% when classifying 4 sleep stages and of 0.48 at an over accuracy of 77.1% when classifying 3 sleep stages (both with TVPP). The table also shows the results obtained without applying TVPP, indicating that using TVPP can help achieve significantly better results. Here the significance was checked with a two-sided

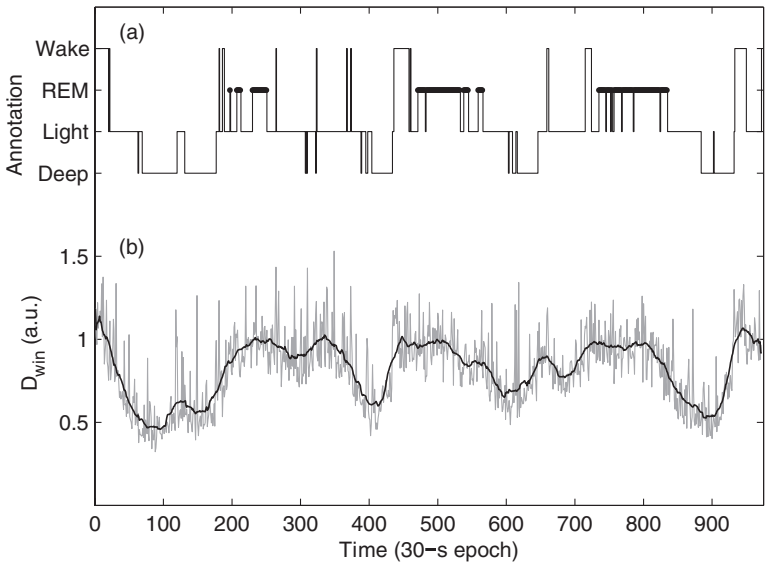

Figure 5. An example of (*a*) overnight annotation and (*b*) feature values of D_{win} from a 50 year-old female, where the unsmoothed (gray) and smoothed (black) feature values are both shown.

Table 3. Discriminative power of D_{win} in separating different sleep stages as evaluated and ranked by ANOVA F-statistic. Results are pooled over all subjects.

Sleep stages	F-statistic	Rank[a]	Top 10 features[b] (descending order)
Wake/REM	10.7[d]	25	12, 13, 3, 5, 4, 14, 20, 21, 18, 22
Wake/light	1487.5[c]	9	13, 14, 7, 4, 5, 3, 15, 17, 27, 16
Wake/deep	3694.4[c]	2	7, 27, 16, 15, 14, 4, 17, 13, 3, 18
REM/light	1679.0[c]	2	7, 27, 14, 15, 20, 21, 16, 22, 24, 23
REM/deep	4915.8[c]	2	7, 27, 16, 20, 15, 21, 25, 22, 23, 24
Light/deep	1420.9[c]	4	16, 15, 7, 27, 17, 14, 10, 4, 8, 18
Wake/REM/light/deep	1912.6[c]	6	7, 16, 15, 13, 14, 27, 4, 5, 17, 3
Wake/REM/NREM	2012.8[c]	4	7, 13, 14, 27, 5, 4, 15, 16, 3, 12

[a] Ranking of F-statistic among all respiratory features.
[b] The feature indices are referred to table 2 and the new feature (feature 27) is indicated with underline.
[c] $p < 0.0001$,
[d] $p < 0.005$.

Wilcoxon signed-rank test. To understand what aspects of sleep staging the new feature improves, we present the confusion matrices obtained with and without D_{win} in table 5 (for 4-stage classification) and in table 6 (for 3-stage classification), where TVPP was applied.

4. Discussion

The deployment of respiratory effort dissimilarity with several consecutive breaths (as measured by a uniform scaling distance) to characterize the regulation of breathing within

Table 4. Ten-fold CV results of 4-stage (wake, REM sleep, light sleep and deep sleep) and 3-stage (wake, REM sleep and NREM sleep) classification schemes obtained using the feature set with and without D_{win}, where the results obtained with and without using TVPP are also presented.

		Without D_{win} [a]		With D_{win} [b]	
Scheme	TVPP	Accuracy	Kappa (κ)	Accuracy	Kappa (κ)
4 stages	No	53.7 ± 8.3%	0.34 ± 0.12	55.2 ± 8.0% [c]	0.37 ± 0.11[c]
	Yes	63.8 ± 8.0%	0.38 ± 0.14	64.9 ± 7.8% [d]	0.41 ± 0.14[c]
3 stages	No	69.2 ± 9.7%	0.43 ± 0.16	70.0 ± 9.3% [d]	0.45 ± 0.15[d]
	Yes	76.1 ± 7.8%	0.45 ± 0.16	77.1 ± 7.6% [d]	0.48 ± 0.17[c]

[a] 26 existing features
[b] 27 features (26 existing features and D_{win}).
Note: Significance of difference was found with and without D_{win} using a paired Wilcoxon signed-rank test (two-sided) at [c]$p<0.001$, or [d]$p<0.01$.

Table 5. Confusion matrix of 4-stage classification (ten-fold CV) obtained using feature set with and without D_{win}, where the results without D_{win} are given in parentheses.

PSG ↓ Classified →	Wake	REM sleep	Light sleep	Deep sleep
Wake	2608 (2606)	512 (453)	2533 (2622)	56 (28)
REM sleep	269 (288)	4259 (3679)	3992 (4492)	13 (74)
Light sleep	844 (831)	2018 (1839)	19285 (19569)	1883 (1791)
Deep sleep	35 (33)	55 (65)	3532 (3664)	2887 (2747)

Table 6. Confusion matrix of 3-stage classification (ten-fold CV) obtained using feature set with and without D_{win}, where the results without D_{win} are given in parentheses.

PSG ↓ Classified →	Wake	REM sleep	NREM sleep
Wake	2605 (2596)	540 (495)	2564 (2618)
REM sleep	271 (278)	4255 (3909)	4007 (4346)
NREM sleep	851 (861)	2112 (2050)	27576 (27628)

wscore between two deep sleep epochs. This is because respiratory effort during NREM sleep (in particular during deep sleep) is steadier and more regular compared with that during wake and REM sleep as mentioned before. As illustrated in figure 3, the discrimination between wake and REM sleep in terms of respiratory effort dissimilarity over time difference is not consistent and seems maximized at $|\Delta|$ beyond 40 epochs. With smaller time differences, overlap can be observed between the dissimilarity scores for wake–wake and REM–REM comparisons. During wake, breathing control might be somewhat less affected by conscious control as well as body movements or other external influences in a short range (e.g. with a $|\Delta|$ of less than 10 epochs or 5 min). This would decrease the dissimilarity scores of wake–wake comparison during that range, yielding a difficulty in distinguishing between wake and REM sleep. As a result of that, the windowed dissimilarity feature D_{win} has a low discriminative power in separating wake and REM sleep as shown in table 3.

Actually, classifying wake and REM sleep might sometimes be difficult even with PSG-based visual scoring (Silber *et al* 2007).

In this work, we chose the window size w of 25 epochs to compute D_{win} by globally maximizing the feature discriminative power in classifying wake, REM sleep, light sleep and deep sleep. However, it might not be the optimal choice all the time, particularly in separating wake and REM sleep (see figure 3). The optimal window size might vary when classifying different sleep stages. Therefore, we think that using an adaptive window size to discriminate between different sleep stages merits further investigation.

Regarding sleep staging, the new feature D_{win} helped improve the classification performance (table 4) and it contributed more to the detection of REM and deep sleep from the other sleep stages (table 5). It is therefore suggested that this feature contains additional information that is not carried by the existing features. We also reveal that using TVPP can lead to better classification results, as shown in table 4. With cardiorespiratory activity, a κ of 0.46 and an overall accuracy of 76.1% were achieved when classifying wake, REM sleep and NREM sleep for 31 healthy subjects (Redmond *et al* 2007). We obtained slightly better results with the use of the respiratory information alone. For 4-stage classification (wake, REM sleep, light sleep and deep sleep), a κ of 0.48 and an overall accuracy of 65.4% (re-computed based on the reported confusion matrix) were achieved by Hedner *et al* (2011), which outperform our results. However, they employed more signal modalities including peripheral arterial tone, pulse rate, oxyhemoglobin saturation and actigraphy. In a more recent study, Willemen *et al* (2014) reported a κ of 0.56 (at an accuracy of 69%) for 4-stage classification using cardiorespiratory and body movement features, whereas they considered an epoch of 60 s instead of the standard 30 s used in most studies with respect to sleep staging. Nevertheless, we anticipate that combining respiratory and cardiac activity will result in a performance enhancement on sleep stage classification and this will be further studied.

The PSG-based sleep stages were manually scored based on the R&K rules in the SIESTA database. However, it has been reported that the overall inter-scorer agreement using the new AASM standard is slightly higher than that obtained using the R&K rules (Danker-Hopfe *et al* 2009). Therefore, the AASM standard is suggested to be applied for PSG-based sleep stage scoring in future work, which is expected to deliver more reliable annotations of overnight sleep stages used for the task of respiratory-based sleep stage classification.

This study only considered healthy subjects without any reported medical, neurological, mental, or cardiovascular diseases as mentioned before. However, for patients with sleep-disordered breathing (e.g. sleep apnea/hypopnea) or other respiratory abnormalities, abnormal respiratory events during the night can affect measuring the dissimilarity between respiratory effort signals. Therefore, the approach described in this work needs to be tested further for these patients. In addition, it has been shown that the respiratory effort is more sensitive to changes of sleep posture and body movements during sleep in comparison with measurements by nasal cannulas (Whyte *et al* 1991). In that case, D_{win} might be erroneously calculated, thus harming the classification performance. However, for the dissimilarity measure described in this paper, the effect of sleep posture might be eliminated since it was computed by comparing each respiratory signal segment with its adjacent segments where the same sleep posture was expected. Moreover, the dissimilarity measure focused on comparing signal morphology with a certain number of breaths, where the falsely detected peaks and troughs (often corresponding to body movements) were removed. As a result, the influences of sleep posture and body movements should be diminished to some extent. Despite that, those influences merit further investigation.

5. Conclusion

By analyzing continuous overnight respiratory effort from healthy subjects, we found that sleep stages can be differentiated using a dissimilarity measure. This measure expresses the dissimilarity between respiratory effort signals in their morphology. The dissimilarity can be evoked by autonomic activity, alternation of ventilation control, or other external factors. A new feature was extracted based on the properties of respiratory effort dissimilarity. It performed worse than an existing feature (standard deviation of respiratory frequency). However, when combined with all 26 existing respiratory features, the new feature can help improving the performance of sleep staging (except for detecting wake from REM sleep). This indicates that this new feature contains additional information that is not carried by the existing features for sleep staging.

Acknowledgments

The authors would like to thank two anonymous reviewers and Dr L Atallah from Philips Research for their insightful comments.

References

Anderer P *et al* 2005 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* **51** 115–33

Bashan A, Bartsch R P, Kantelhardt J W and Havlin S 2008 Comparison of detrending methods for fluctuation analysis *Physica* A **387** 5080–90

Berry R B *et al* 2012 Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events *J. Clin. Sleep Med.* **8** 597–619

Bresler M, Sheffy K, Pillar G, Preiszler M and Herscovici S 2008 Differentiating between light and deep sleep stages using an ambulatory device based on peripheral arterial tonometry *Physiol. Meas.* **29** 571–84

Buysse D J, Reynolds C F, Monk T H, Berman S R and Kupfer D J 1989 The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research *Psychiatry Res.* **28** 193–213

Cherniack N S 1981 Respiratory dysrhythmias during sleep *New Engl. J. Med.* **305** 325–30

Chung G S, Choi B H, Lee J S, Lee J S, Jeong D U and Park K S 2009 REM sleep estimation only using respiratory dynamics *Physiol. Meas.* **30** 1327–40

Cohen J A 1960 A coefficient of agreement for nominal scales *Edu. Psychol. Meas.* **20** 37–46

Costa M, Goldberger A L and Peng C K 2005 Multiscale entropy analysis of biological signals *Phys. Rev.* E **71** 021906

Danker-Hopfe H *et al* 2009 Interrater reliability for sleep scoring according to the Rechtschaffen and Kales and the new AASM standard *J. Sleep. Res.* **18** 74–84

Devot S, Dratwa R and Naujokat E 2010 Sleep/wake detection based on cardiorespiratory signals and actigraphy *IEEE Conf. Proc. Eng. Med. Bio. Soc. (Buenos Aires, 31 August–4 September 2010)* pp 5089–92

Douglas N J, White D P, Pickett C K, Weil J V and Zwillich C W 1982 Respiration during sleep in normal man *Thorax* **37** 840–4

Foussier J, Fonseca P, Long X and Leonhardt S 2013 Automatic feature selection for sleep/wake classification with small data sets *Int. Joint Conf. Biomed. Eng. Sys. Technol. (Barcelona, Spain, 11–14 February 2013)* pp 178–84

Hedner J, White D P, Malhotra A, Herscovici S, Pittman S D, Zou D, Grote L and Pillar G 2011 Sleep staging based on autonomic signals: a multi-center validation study *J. Clin. Sleep Med.* **7** 301–6

Heinzer R C and Sériès F 2011 Normal physiology of the upper and lower airways *Principles and Practice of Sleep Medicine* Ed M H Kryger, T Roth and W C Dement 5th edn (Amsterdam: Elsevier) chapter 23

Herscovici S, Peer A, Papyan S and Lavie P 2007 Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone and actigraphy *Physiol. Meas.* **28** 129–40

Horvatic D, Stanley H E and Podobnik B 2011 Detrended cross-correlation analysis for non-stationary time series with periodic trends *Europhys. Lett.* **94** 18007

Iber C, Ancoli-Israel S and Chesson A L 2007 *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* (Westchester, IL: American Academy of Sleep Medicine)

Kantelhardt J W, Penzel T, Rostig S, Becker H F, Havlin S and Bunde A 2003 Breathing during REM and non-REM sleep: correlated versus uncorrelated behaviour *Physica* A **319** 447–57

Klösch G *et al* 2001 The SIESTA project polygraphic and clinical database *IEEE Eng. Med. Biol. Mag.* **20** 51–7

Kortelainen J M, Mendez M O, Bianchi A M, Matteucci M and Cerutti S 2010 Sleep staging based on signals acquired through bed sensor *IEEE Trans. Inform. Technol. Biomed.* **14** 776–85

Lázaro J, Gil E, Bailón R, Minchole A and Laguna P 2013 Deriving respiration from photoplethysmographic pulse width *Med. Biol. Eng. Comput.* **51** 233–42

Long X, Fonseca P, Foussier J, Haakma R and Aarts R M 2014a Sleep and wake classification with actigraphy and respiratory effort using dynamic warping *IEEE J. Biomed. Health Inform.* **18** 1272–84

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

Matthews G, Sudduth B and Burrow M 2000 A non-contact vital signs monitor *Crit. Rev. Biomed. Eng.* **28** 173–8

Ohayon M M, Carskadon M A, Guilleminault C and Vitiello M V 2004 Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan *Sleep* **27** 1255–73

Penzel T, Wessel N, Riedl M, Kantelhardt J W, Rostig S, Glos M, Suhrbier A, Malberg H and Fietze I 2007 Cardiovascular and respiratory dynamics during normal and pathological sleep *Chaos* **17** 015116

Phillipson E A 1978 Control of breathing during sleep *Am. Rev. Respir. Dis.* **118** 909–39

Polkey M I, Green M and Moxham J 1995 Measurement of respiratory muscle strength *Thorax* **50** 1131–5

Rechtschaffen E A and Kales A 1968 *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* (Washington, DC: Bethesda)

Redmond S J, de Chazal P, O'Brien C, Ryan S, McNicholas W T and Heneghan C 2007 Sleep staging using cardio-respiratory signals *Somnologie* **11** 245–56

Redmond S J and Heneghan C 2006 Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea *IEEE Trans. Biomed. Eng.* **53** 485–96

Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy *Am. J. Physiol.* **278** H2039–49

Roebuck A, Monasterio V, Gederi E, Osipov M, Behar J, Malhotra A, Penzel T and Clifford G D 2014 A review of signals used in sleep analysis *Physiol. Meas.* **35** R1–57

Rostig S, Kantelhardt J W, Penzel T, Cassel W, Peter J H, Vogelmeier C, Becker H F and Jerrentrup A 2005 Nonrandom variability of respiration during sleep in healthy humans *Sleep* **28** 411–7

Samy L, Huang M C, Liu J, Xu W and Sarrafzadeh M 2014 Unobtrusive sleep stage identification using a pressure-sensitive bed sheet *IEEE Sens.* J. **14** 2092–101

Schumann A Y, Bartsch R P, Penzel T, Ivanov P and Kantelhardt J W 2010 Aging effects on cardiac and respiratory dynamics in healthy subjects across sleep stages *Sleep* **33** 943–55

Silber M H *et al* 2007 The visual scoring of sleep in adults *J. Clin. Sleep Med.* **3** 121–31

Somers V K, Dyken M E, Mark A L and Abboud F M 1993 Sympathetic-nerve activity during sleep in normal subjects *New Engl. J. Med.* **328** 303–7

Trinder J, Kleiman J, Carrington M, Smith S, Breen S, Tan N and Kim Y 2001 Autonomic activity during human sleep as a function of time and sleep stage *J. Sleep Res.* **10** 253–64

Watanabe K, Watanabe T, Watanabe H, Ando H, Ishikawa T and Kobayashi K 2005 Noninvasive measurement of heartbeat, respiration, snoring and body movements of a subject in bed via a pneumatic method *IEEE Trans. Biomed. Eng.* **52** 2100–7

Whyte K F, Gugger M, Gould G A, Molloy J, Wraith P K and Douglas N J 1991 Accuracy of respiratory inductive plethysmograph in measuring tidal volume during sleep *J. Appl. Physiol.* **71** 1866–71

Willemen T, Van Deun D, Verhaert V, Vandekerckhove M, Exadaktylos V, Verbraecken J, Van Huffel S, Haex B and Vander Sloten J 2014 An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification *IEEE J. Biomed. Health Inform.* **18** 661–9

Yankov D, Keogh E, Medina J, Chiu B and Zordan V 2007 Detecting time series motifs under uniform scaling *ACM Conf. Proc. ACM SIGKDD (San Jose, CA, USA, 12–15 August 2007)* pp 844–53