Detection of Nocturnal Slow Wave Sleep Based on Cardiorespiratory Activity in Healthy Adults

Xi Long, *Member, IEEE*, Pedro Fonseca, Ronald M. Aarts, *Fellow, IEEE*, Reinder Haakma, Jérôme Rolink, *Member, IEEE*, and Steffen Leonhardt, *Senior Member, IEEE*

***Abstract*—Human slow wave sleep (SWS) during bedtime is paramount for energy conservation and memory consolidation. This study aims at automatically detecting SWS from nocturnal sleep using cardiorespiratory signals that can be acquired with unobtrusive sensors in a home-based scenario. From the signals, time-dependent features are extracted for continuous 30-s epochs. To reduce the measuring noise, body motion artifacts, and/or within-subject variability in physiology conveyed by the features, and thus, enhance the detection performance, we propose to smooth the features over each night using a spline fitting method. In addition, it was found that the changes in cardiorespiratory activity precede the transitions between SWS and the other sleep stages (non-SWS). To this matter, a novel scheme is proposed that performs the SWS detection for each epoch using the feature values prior to that epoch. Experiments were conducted with a large dataset of 325 overnight polysomnography (PSG) recordings using a linear discriminant classifier and tenfold cross validation. Features were selected with a correlation-based method. Results show that the performance in classifying SWS and non-SWS can be significantly improved when smoothing the features and using the preceding feature values of 5-min earlier. We achieved a Cohen's Kappa coefficient of 0.57 (at an accuracy of 88.8%) using only six selected features for 257 recordings with a minimum of 30-min overnight SWS that were considered representative of their habitual sleeping pattern at home. These features included the standard deviation, low-frequency spectral power, and detrended fluctuation of heartbeat intervals as well as the variations of respiratory frequency and upper and lower respiratory envelopes. A marked drop in Kappa to 0.21 was observed for the other nights with SWS time of less than 30 min, which were found to more likely occur in elderly. This will be the future challenge in cardiorespiratory-based SWS detection.**

***Index Terms*—Cardiorespiratory activity, feature smoothing, slow wave sleep detection, spline fitting, time delay.**

Manuscript received March 2, 2015; revised July 28, 2015 and September 2, 2015; accepted September 29, 2015. Date of publication October 6, 2015; date of current version January 31, 2017.

X. Long, P. Fonseca, and R. M. Aarts are with the Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands and also with the Philips Research, High Tech Campus, 5656 AE Eindhoven, The Netherlands (e-mail: x.long@tue.nl; pedro.fonseca@philips.com; ronald.m.aarts@philips.com).

R. Haakma is with the Philips Research, High Tech Campus, 5656 AE Eindhoven, The Netherlands (e-mail: reinder.haakma@philips.com).

J. Rolink and S. Leonhardt are with the Philips Chair for Medical Information Technology, RWTH Aachen University, D-52074, Aachen, Germany (e-mail: rolink@hia.rwth-aachen.de; leonhardt@hia.rwth-aachen.de).

Digital Object Identifier 10.1109/JBHI.2015.2487446

I. Introduction

NOCTURNAL sleep of humans is comprised of rapid-eye-movement (REM) sleep, stages S1–S4 of non-REM (NREM) sleep, and wake according to the R&K rules [1]. S1 and S2 are grouped into "light sleep," where S1 and S2 correspond to stages N1 and N2, respectively, according to the more recent guidelines of the American Academy of Sleep Medicine (AASM) [2]. S3 and S4 are considered slow wave sleep (SWS), in correspondence to N3 stage in the AASM guidelines. SWS relates to delta electroencephalographic (EEG) activity with no eye movements [2]. It represents the most restorative period of sleep for metabolic functioning, during which brain and body energies are conserved [3] and new memories are consolidated [4]. Thus, SWS is associated with maintenance of sleep and sleep quality [5]. Lack of SWS may result in, e.g., loss of daytime performance [5] and increased risk of diabetes [6]. More interestingly, attention has been engaged in the past decade to improve nighttime sleep (e.g., to enhance memory consolidation) through external stimulation of sleep slow waves in humans [7]–[9]. Therefore, we were motivated to develop a system to accurately detect SWS from nocturnal sleep, particularly in a home scenario.

Polysomnography (PSG) is the "gold standard" for objective sleep assessment, relying on which a hypnogram can be derived through visual scoring by sleep technicians [1], [2]. A PSG recording typically consists of various biosignals such as EEG, electromyograhy (EMG), electrooculography (EOG), electrocardiography (ECG), respiratory effort (RE), and blood oxygen saturation (SpO2). These signals are usually split into continuous 30-s nonoverlapping intervals, called epochs. Although PSG is a standard method for sleep analysis, it has some disadvantages, for example, it is conducted in a sleep laboratory, leading to high costs with facilities; it requires many electrodes to be attached to the body, disrupting a subject's normal sleep as a consequence; and it requires subjects to stay in the sleep laboratory overnight that is not compatible with a prolonged sleep monitoring. To overcome these disadvantages, cardiac/respiratory information has been deployed to assess sleep for years as long as they can be acquired with unobtrusive sensing systems in a home-based environment such as with a wrist-worn watch [10], a bed sensor [11], a textile bedsheet [12], a web camera [13], an acoustic device [14], a Doppler radar [15], and a photoplethysmographic sensor [16]. It has been proven that cardiorespiratory signals contain relevant physiological information for sleep staging such as heart rate variability (HRV) [17] and respiration

rhythm [18]. This is because they are related to autonomic nervous system that differs between sleep stages [19], [20]. For example, SWS coincides with an decreased sympathetic activity conveyed by the low-frequency power in HRV [19], [21].

Cardiorespiratory-based sleep stage classification has been increasingly studied in recent years, where many features (representing certain physiological aspects) have been designed and extracted from cardiac and/or respiratory signals [22]–[26]. However, rather than SWS detection, those studies investigated either wake–REM–NREM or sleep–wake classification. Many other studies have reported results in classifying wake, REM sleep, light sleep, and SWS [10], detecting REM sleep [27], or differentiating light sleep and SWS [28], whereas they used additional physiological signal modalities such as peripheral arterial tone and oxyhemoglobin saturation. Shinar *et al.* [29] developed an HRV-based SWS detector and obtained an accuracy of about 80%, while they used a very small portion with a total duration of 100 min (SWS of 50 min) rather than entire-night recordings for validation.

The sleeping pattern of healthy adults usually progresses with several regular cycles throughout the night [30]. This means that, for each recording, the sleep stage with associated physiological activity across the night is time variant so that each feature is considered an epoch-based time series. After visually comparing some feature values and PSG-based annotations changing over night, we observed many errors occurring in the middle of a long SWS or non-SWS (all the other stages) period, possibly due to measuring noise, feature computing variances, or body motion artifacts. Another cause might be the "within-subject variability" in physiology, meaning that the physiological expression of features was not perfectly discriminative, and thus, could not deliver an ideal separation between sleep stages. For these reasons, we decided to low-pass filter or smooth each feature's values over time using a spline fitting method [31]. The main reason of using spline fitting was that it is capable of interpolating missing data compared with many other low-pass filters [32]. This is of particular importance because sleep is a continuous process and we found that our data had an average of $\sim$10% missing values.

Several researchers have investigated the temporal relationship between cardiac dynamics and brain activity [33]–[35]. For instance, Otzenberger *et al*. [33] reported that the overnight HRV changes generally precede the variations in EEG activity by around 1–2 min. Jurysta *et al.* [34] demonstrated that the high-frequency power of heartbeat or RR intervals corresponds to a preceding time (or negative time delay) of several minutes compared with the delta-wave power of EEG spectrum. More specifically, our recent work [35] showed that, during the transitions between SWS and non-SWS, the changes in cardiac activity anticipate the variations in EEG activity by about 2 min. These studies indicate that the autonomic changes are not exactly synchronized with the EEG variations for the transitions between SWS and non-SWS; rather that a time difference appears in between. This time-delay phenomenon would end up with errors in SWS detection. To this matter, we propose a novel scheme by using the preceding feature values in earlier epochs to further improve the identification of the sleep state of each epoch (SWS or non-SWS).

This paper addresses the problem of continuously classifying overnight SWS and non-SWS with cardiorespiratory signals that can be unobtrusively acquired. Previous work has shown that a linear discriminant (LD) classifier is appropriate in the problem of sleep stage classification [22], [25], [36], which was adopted in this study for SWS and non-SWS classification. Preliminary results of this work have been previously reported [37].

II. Subjects and Data

Full PSG data (at least 16 channels of biosignals) from 165 healthy subjects in the SIESTA project [38] was included, monitored in seven different sleep centers located in five European countries. In accordance with the SIESTA protocol, the subjects met several criteria such as no reported symptoms of neurological, mental, medical, or cardiovascular disorders, no history of drug or alcohol abuse, no shift work, and retirement to bed before midnight depending on their habitual bedtime [38]. Each subject spent two consecutive nights in a sleep laboratory, resulting in a total of 330 overnight recordings. For each recording, the scoring of 30-s epoch-based sleep stages was carried out by sleep technicians based on PSG according to the R&K rules. For SWS and non-SWS classification, wake, REM sleep, S1, and S2 were merged into a single *non-SWS* class; S3 and S4 were labeled as *SWS* class. The epochs with invalid PSG scoring ($\sim$3%) were removed.

Five recordings were excluded due to the absence of SWS, yielding an inclusion of 325 recordings in our dataset. In addition, this study primarily addressed on the "normal" sleep nights (from lights OFF in the evening till lights ON in the morning), during which the total SWS time throughout the night was no usually less than 30 min found in [39], resulting a group of 257 recordings from 145 subjects in a normal group. These nights were more representative of the normal sleeping pattern in terms of SWS [39], which were expected with a home-based sleep monitoring. Carskadon and Dement [30] reported that SWS generally constitutes approximately 13–23% of sleep for healthy subjects, confirming the validity of our selection of recordings. The remaining 68 nights (from 51 subjects) with the overnight total SWS time of less than 30 min (low-SWS group), more from the first nights ($\sim$60%) than the second nights, were excluded because they might be strongly influenced by the "laboratory effects," where the subjects could not sleep well as habitual as being at home [40]. The subject demographics and sleep data for the normal group used in this study is summarized in Table I. In spite of that, we also tested our approach on the recordings from the low-SWS group.

The thoracic RE signals (sampled at 10 Hz) were acquired with a respiratory inductance plethysmographic chest belt and the cardiac signals (sampled at $\geq$100 Hz) were recorded with a modified V1 lead ECG.

III. Methods

A. Signal Preprocessing

The RE signal was filtered with a tenth-order Butterworth low-pass filter (with a cutoff frequency of 0.6 Hz) to eliminate high-frequency noise; afterwards, the baseline was subtracted

TABLE I
SUBJECT DEMOGRAPHICS AND SLEEP DATA FROM NORMAL NIGHTS (WITH A MINIMAL SWS TIME OF 30 MIN) FOCUSED IN THIS STUDY

Parameter	Mean ± Std	Range
Number of recordings	$N = 257$ (145 subjects)	
Sex	65 males and 80 females	
Age (year)	49.5 ± 19.2	20 – 95
Body mass index, BMI (kg/m^2)	24.3 ± 3.4	17.0 – 34.8
Total recording time (h)	7.8 ± 0.5	5.7 – 9.3
Wake (%)	17.5 ± 11.0	1.1 – 63.0
REM (%)	15.8 ± 5.5	0.0 – 29.0
Light (%)	51.9 ± 8.6	21.1 – 70.4
SWS (%)	14.8 ± 5.1	6.2 – 32.2

by the median peak-to-trough amplitude over the entire recording [25], [41]. Because we also extracted respiratory features in the frequency domain, a fast Fourier transform with a Hanning window (used to reduce spectral leakage) was applied to estimate the power spectral density (PSD) on the resulting signal for each epoch [41]. The body motion artifacts were not excluded when computing RE features since they were expected to be indicative of wake epochs to a certain extent, during which body movements often occur.

The ECG signal was high-pass filtered using a Kaiser window (with a cutoff frequency of 0.8 Hz and a side-lobe attenuation of 30 dB) to remove baseline wander [42], after which the resulting signal was zero meaned. To extract features from RR intervals for each epoch, a Hamilton–Tompkins R-peak detector [43] combined with a precise QRS localization algorithm [44] was applied to locate R peaks on the ECG signal with a window of nine epochs centered at the epoch of interest. This window served to include sufficient data points to capture the changes in RR intervals, where the window size is close to the value of 5 min recommended in [45]. The R-peak localization algorithm developed in our previous work [44] has been shown to be robust to motion artifacts, where ectopic RR intervals longer than 2 s or shorter than 0.3 s were excluded. The resulting RR interval series was then resampled via linear interpolation at a sampling rate of 4 Hz. The PSD of RR intervals was estimated using an autoregressive (AR) model with adaptive order [46]. Here, the AR model instead of a Fourier-based approach was implemented due to its limitations such as poor spectral resolution and leakage [47], which were more sensitive to estimating the PSD of the RR interval series having a lower sampling rate compared with the RE signal.

B. Feature Extraction

A total of 70 features were extracted for each 30-s epoch from ECG and thoracic RE signals, which are briefly described later. Note that the features for a specific epoch were mostly computed within a certain window centered at that epoch.

The ECG features were obtained from the RR intervals or heart rates. In the time domain, they included the mean heart rate, mean RR interval (detrended and non-detrended), standard deviation and range of RR intervals, the percentage of successive RR intervals that differ by more than 50 ms, and the root mean square and standard deviation of successive RR interval differences [22], [45]. These features were computed over a window of nine epochs (with around 300 beats during sleep on average), where the window size was the same as that used for estimating RR PSD, as recommended in [45]. Frequency domain features comprised the logarithm of normalized power in the very low frequency (VLF, 0.003-0.04 Hz), low frequency (LF, 0.04–0.15 Hz), and high frequency (HF, 0.15–0.4 Hz) spectral bands, the ratio of LF and HF spectral powers [22], [45], and the module and phase of HF pole [24]. The VLF, LF, and HF power and LF-to-HF ratio with adapted spectral bands have succeeded in improving sleep/wake detection [48]. The maximum power in the HF band and its associated frequency (in line with the mean respiratory frequency) were also calculated [41]. Additionally, nonlinear properties of RR intervals were quantified based on detrended fluctuation analysis (DFA) with parameter α [49] and its short-term (parameter α_1) and long-term (parameter α_2) exponents [50], and multiscale sample entropy (length: 1 and 2 samples, scale: 1–10) [51].

The RE features included the mean respiratory frequency estimated in the time and the frequency domain, respiratory frequency standard deviation over 150 s, mean and standard deviation of breath-by-breath correlations, standard deviation of breath lengths, and the spectral power of respiratory frequency [22], [41]. Several features regarding the RE amplitude were derived: the standardized median and sample entropy of respiratory peaks and troughs (i.e., the respiratory upper and lower envelopes indicating the inhalation and exhalation depths, respectively), median peak-to-trough difference, median volumes and flow rates of breath cycles, inhalations, and exhalations, and the ratio of inhalation and exhalation flow rates [25]. When computing these amplitude-based features for each epoch, we used a window of 13 epochs (with around 120 breath cycles) since the sample entropy measures are less reliable if the number of samples is less than 100 [52]. From the spectrum, we found the power in different spectral bands (VLF: 0.01–0.05 Hz, LF: 0.05–0.15 Hz, and HF: 0.15–0.5 Hz) and the LF-to-HF ratio obtained from respiratory PSD as suggested by Redmond *et al.* [22], albeit there is currently no standard of analyzing the respiratory power spectrum. In addition, we extracted the respiratory regularity quantified with sample entropy over seven epochs [25] and windowed respiratory dissimilarity measured by means of uniform scaling [53] and dynamic (time and frequency) warping [36], respectively.

All the features were expected to associate with sleep stages but it is unclear which of them are more indicative in identifying SWS. Besides, many of them might be mutually correlated containing redundant information for SWS detection. Thus, a feature selection algorithm taking into account the feature discriminative power as well as feature correlation should be applied, which will be explained in Section III-D.

C. Spline Fitting for Feature Smoothing

As stated, the features should cycle with time in terms of sleep stage, which motivated us to consider a recording- or night-specific feature smoothing. Before that, each feature was normalized for each recording to have zero mean and unit variance (Z-score normalization). This served to reduce the variability between subjects caused by the difference between

PSG systems used in different sleep laboratories and/or the difference in physiological expression during sleep. Our previous work [37] has revealed that the Z-score normalization can help improving SWS detection.

The spline fitting method has been widely used for time-series smoothing [32]. Let $\mathbf{x}$ represent a sequence of observations $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ $(x_1 < x_2 < ... < x_n)$ and $\mathbf{y}$ their responses $\mathbf{y} = \{y_1, y_2, ..., y_n\}$, then a relation between them can be modeled by

$$y_i = g(x_i) + \varepsilon_i \ (i = 1, 2, ..., n) \tag{1}$$

where g is a smoothing (spline) function and ε_i are independent and identically distributed residuals. The smoothing function can be estimated by minimizing the objective function (i.e., the penalized sum of squares) such that

$$\hat{g} = \arg\min_{g} \left[\sum_{i=1}^{n} [y_i - g(x_i)]^2 + \lambda \int_{x_1}^{x_n} g''(x)^2 dx \right] \tag{2}$$

where λ is a smoothing parameter that controls the tradeoff between residual and local variation. The smoothing function can be expressed by cubic B-splines as basis functions and determined via least-squares approximation (LSA) [31], [32].

Given a feature for a specific recording, the observations here are the epoch indices $\mathbf{t} = \{t_1, t_2, ..., t_m\}$ and the responses are their corresponding feature values $\mathbf{v} = \{v_1, v_2, ..., v_m\}$, where m is the total number of epochs. To build up a spline fitting model, the entire sequence is divided in k continuous subsequences with $k-1$ boundaries called knots or breaks; and each of them contains l epochs. The feature values and epoch indices for this recording are then expressed, respectively, as

$$\mathbf{v} = \{\underbrace{v_{11}, v_{12}, ..., v_{1l}}_{1}, \underbrace{v_{21}, v_{22}, ..., v_{2l}}_{2}, ..., \underbrace{v_{k1}, v_{k2}, ..., v_{kl}}_{k}\} \tag{3}$$

and

$$\mathbf{t} = \{\underbrace{t_{11}, t_{12}, ..., t_{1l}}_{1}, \underbrace{t_{21}, t_{22}, ..., t_{2l}}_{2}, ..., \underbrace{t_{k1}, t_{k2}, ..., t_{kl}}_{k}\}. \tag{4}$$

Thereafter, each subsequence is modeled by (1) and (2), yielding a spline fitting over the entire sequence with multiple knots. Since the total number of epochs differed between recordings, we preferred to fix the window size of subsequences $w = \lceil m/k \rceil$ instead of using a fixed number of breaks k. A larger window size (or fewer knots) results in a smoother fitting curve; while a smaller window size (or more knots) decreases its smoothness. For example, as depicted in Fig. 1, the feature values throughout the night after spline smoothing seem better mapped to the PSG-based annotations. The figure also shows that the RR interval and respiratory frequency had lower variances during SWS compared with the other stages.

D. Feature Subset Selection

Since an LD classifier is usually sensitive to the presence of redundant and nondiscriminative features, classification performance would degrade as a result. Hence, we applied a correlation-based feature selector (CFS) [54] aiming at selecting an "optimal" subset of features that are highly correlated with the classification (high feature discriminative power) but uncorrelated to each other (low feature redundancy). With the CFS, the heuristic evaluation criterion, called "merit," can be formulized by taking the feature-to-class and feature-to-feature correlations into account. Starting with no features, a forward search was used to add new features one-by-one until no increase on merit was observed when in combination with additional features. More details of the CFS can be found elsewhere [54].

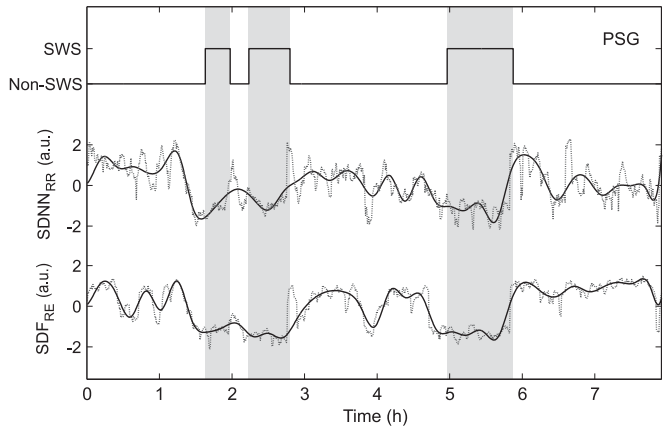


Fig. 1. Example of overnight PSG-based annotations of SWS and non-SWS and the values of two representative features $\mathrm{SDNN_{RR}}$ (standard deviation of RR intervals) and $\mathrm{SDF_{RE}}$ (standard deviation of respiratory frequency) from a subject. The unsmoothed (dashed) and smoothed (solid) feature values are plotted. The window width for spline fitting was 25 epochs. By comparing the annotations and the two features, classification errors might occur around the transitions between SWS and non-SWS (e.g., the transition around the fifth hour).

E. Classifier

Here, a simple LD classifier was adopted and the classification was performed on each epoch over the whole recording. The LD function is given by

$$G_c(\mathbf{f}) = -\frac{1}{2}(\mathbf{f} - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}^{-1}(\mathbf{f} - \boldsymbol{\mu}_c) + \ln Pr(c) \tag{5}$$

where $\boldsymbol{\mu}_c$ expresses the mean of the feature vector $\mathbf{f}$, $\boldsymbol{\Sigma}$ the pooled covariance matrix, and $Pr(c)$ the prior probability for class c [*SWS* (positive class) or *non-SWS* (negative class)]. Given a feature vector, the jth epoch E_j $(j = 1, 2, ..., m)$ of a recording can be classified using the decision making rule

$$\mathcal{C}(E_j \mid \mathbf{f}_j) = \begin{cases} SWS & \text{if } G_{SWS}(\mathbf{f}_j) > G_{non-SWS}(\mathbf{f}_j) \\ non\text{-}SWS & \text{otherwise.} \end{cases} \tag{6}$$

We observed that the occurrence of each class varied throughout the night. For instance, the probability of being in SWS at the end of the night should be lower than that in the middle of the night. This indicates that the prior probabilities are time varying. Instead of using a fixed prior probability, we used the time-varying prior probability $\ln Pr(c, t)$ for class c (*SWS* or *non-SWS*) at time t since lights OFF (i.e., epoch index). For each epoch at time t, it was computed by simply counting the

relative frequency where the epochs (from all other training recordings) in that specific time of night t were annotated as each class (*SWS* and *non-SWS*) [22], [53].

F. Time Delay

As illustrated in Fig. 1, there seems to be some errors in feature values with a few minutes before the transitions between SWS and non-SWS, implying the presence of time delay between the changes of cardiorespiratory properties and the PSG-based annotations. Under the consideration of the time delay, earlier cardiorespiratory activity can be utilized to identify *SWS* or *non-SWS* class. Supposing that we want to classify the jth epoch E_j ($j = 1, 2, ..., m$), we can use the feature values of the (j+τ)th epoch (with a delay of τ epochs to the target epoch) instead of using the feature values from the epoch itself, such that

$$\mathcal{C}(E_j \mid \mathbf{f}_{j+\tau}) = \begin{cases} SWS & \text{if } G_{SWS}(\mathbf{f}_{j+\tau}) \\ & > G_{non\text{-}SWS}(\mathbf{f}_{j+\tau}) \\ non\text{-}SWS & \text{otherwise} \end{cases} \quad (7)$$

in which a negatively delayed time (i.e., a preceding time) was expected. This means that we anticipated the class of the target epoch with τ epochs earlier. To evaluate this approach, we computed the discriminative power of the features and the classification results by varying the time delay from -30 to 0 epochs with a step size of one epoch (a τ of zero corresponds to the absence of time delay).

IV. Experiments and Evaluation

From a practical point of view, we considered a subject-independent cross validation—the two nights' recordings from the same subject were either included in the training or the test dataset. To provide an unbiased evaluation of our classifier, a tenfold cross validation (CV) procedure was conducted. The dataset was partitioned into ten subsets containing recordings as nearly equal as possible. During each iteration of the tenfold CV, nine subsets were used to generate feature subsets, and then, train the classifier and the remaining was used for testing. The classification results were then obtained on each test dataset of the cross validation; thereafter, the evaluation of the classifier's performance was formed by pooling (i.e., aggregating) or averaging all results.

Two rounds of CV were executed in our experiments (one for assembling a unique feature set and the other for generating final classification results). During the first round, to prevent selecting features upon the whole dataset, and thus, biasing the classifier, CFS was applied during each iteration of the tenfold CV, yielding ten "optimal" feature subsets, one for each training set. After that, in order to assemble a single feature list, only the features appearing in all feature subsets were selected. During the second round, this unique list of features was then used in all iterations of tenfold CV to test the classifier.

Although the feature selector can automatically choose features that optimally separate the classes *SWS* and *non-SWS*, evaluating the discriminative power of each single feature explores which physiological aspects help distinguish both classes. It not only allows for the comparison among features but also indicates to what extent the smoothing and time delay help improve the features. For these purposes, the absolute standardized mean difference (ASMD) [55] was employed to measure the discriminative power of a single feature, equivalent to the 1-D Mahalanobis distance [56]. Given a feature f, it is computed as the absolute mean difference of the feature values between *SWS* and *non-SWS* epochs divided by the standard deviation of the values over all epochs

$$\text{ASMD}^{\text{f}} = \frac{|\mu^{\text{f}}_{SWS} - \mu^{\text{f}}_{non-SWS}|}{\sigma^{\text{f}}} \quad (8)$$

where μ^{f}_{SWS} and $\mu^{\text{f}}_{non-SWS}$ express the sample mean of *SWS* and *non-SWS* epochs, respectively, and σ^{f} is the sample standard deviation. A higher discriminative power in separating the two classes translates to a larger ASMD value.

Overall accuracy, precision, sensitivity, and specificity were first considered to evaluate the classifier. However, they might not be appropriate criteria for the "imbalanced class distribution" in our data, where the *non-SWS* epochs account for an average of 87.6% of the night. The Cohen's Kappa coefficient of agreement κ [57] offers an indication of the general classification performance in correctly identifying imbalanced classes by compensating for the probability of chance agreement. Here, the classifier threshold was chosen to optimize the pooled Kappa based on training data. To have an overview of the classification performance across the entire solution space, a precision-recall (PR) curve was used. It plots precision versus recall (or sensitivity) by varying the classifier threshold used to separate the two classes. When comparing classifiers, the metric "area under the PR curve" (AUC_{PR}) was calculated with the algorithm described by Davis and Goadrich [58]. A larger AUC_{PR} corresponds to a better classification performance.

In order to evaluate the effectiveness of the feature smoothing and the time-delay approaches in improving SWS and non-SWS classification, we compared four classification schemes by using features

1) *A*: without smoothing and without time delay,
2) *B:* with smoothing but without time delay,
3) *C:* without smoothing but with time delay, and
4) *D:* with smoothing and time delay.

The spline window size and the delayed time were determined to optimize κ based on training data. Moreover, the classification performance was also compared between using only ECG and only RE signals and between the normal group (the focus of this study) and the low-SWS group.

V. Results

After the feature selection procedure described before, a total of six features (SDNN_{RR}, LF_{RR}, DFA_{RR}, SDF_{RE}, SDMP_{RE}, and SDMT_{RE}) were selected by the CFS when including all cardiorespiratory features. In the same way, we obtained a list of four features (SDNN_{RR}, LF_{RR}, DFA_{RR}, and SE_{RR}) when using ECG alone and four (SDF_{RE}, SDMP_{RE}, SDMT_{RE}, and UNIS_{RE}) when using solely RE. The selected features are listed

TABLE II
FEATURES SELECTED USING CFS FOR SWS DETECTION

Feature Description	Denotation	Modality
RR standard deviation*, †	$SDNN_{RR}$	ECG
RR spectrum power LF band*, †	LF_{RR}	ECG
RR DFA (parameter α)*, †	DFA_{RR}	ECG
RR sample entropy (length 2, scale 1)†	SE_{RR}	ECG
Resp. frequency standard deviation*, ‡	SDF_{RE}	RE
Resp. peak standardized median*, ‡	$SDMP_{RE}$	RE
Resp. trough standardized median*, ‡	$SDMT_{RE}$	RE
Resp. uniform scaling dissimilarity‡	$UNIS_{RE}$	RE

*Six features were selected when using both ECG and RE signals.
†Four features were selected when using only ECG signal.
‡Four features were selected when using only RE signal.

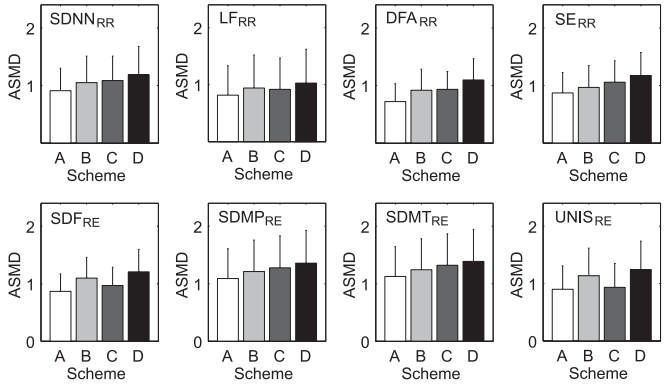


Fig. 2. Average discriminative power (as measured by ASMD) of the selected features in different schemes. The ASMD of scheme D was found to be significantly higher than the others for all the selected features using a paired (two-sided) Wilcoxon signed-rank test ($p < 0.001$). The time delay τ was -2.5 min for scheme C and -5 min for scheme D.

in Table II, where the table footnote indicates the selected features when using different signal modalities.

The averaged discriminative powers of the selected features in different schemes are compared in Fig. 2. It indicates that the smoothing with spline fitting can improve the feature discriminative power. Experimentally it was found that the κ value was maximized at a spline window of 25 epochs. On the other hand, using the features with negative time delay also increased their discriminative power by comparing the ASMD values between schemes A and C (or between schemes B and D). Here, the optimal time delay τ of -2.5 and -5 min were experimentally found for schemes C and D, respectively. As shown in Fig. 3, the selected features were found to significantly differ between SWS and non-SWS.

Fig. 4 plots the classification performance (κ and AUC_{PR}) versus time delay (τ) in schemes C and D. The figure shows that the highest κ and AUC_{PR} occurred with a negative time delay of five epochs (2.5 min) for the unsmoothed features and of ten epochs (5 min) for the smoothed features. This means that the optimal time delay should depend on the window size of spline fitting. As we expected, it was longer in scheme D (with smoothing) than in scheme C (without smoothing).

The results of SWS and non-SWS classification obtained with respect to the four schemes are summarized in Table III. The best result, obtained with smoothing and time delay, corresponds to a

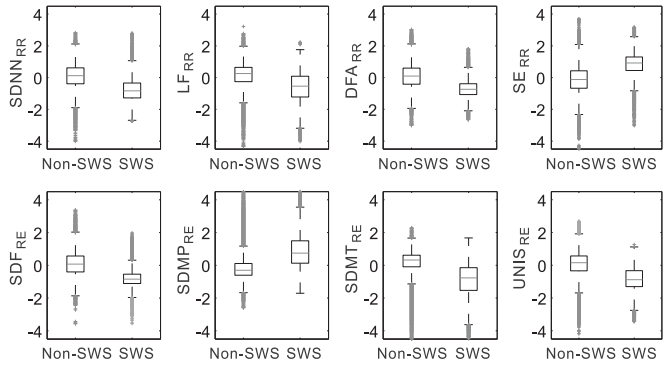


Fig. 3. Boxplots of values of the selected features (in scheme D) in SWS and non-SWS. The significance of difference was found between the two classes for each feature using an unpaired Mann–Whitney test at $p < 0.00001$.

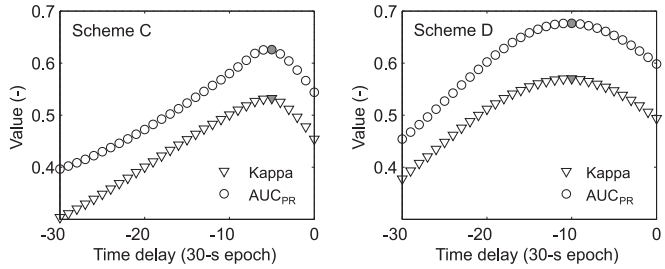


Fig. 4. Classification performance obtained using the six cardiorespiratory features selected via CFS (see Table II) and tenfold CV in scheme C (without smoothing) and scheme D (with smoothing) versus time delay (τ), in epochs. The minus sign of τ indicates the use of preceding feature values.

pooled κ of 0.57, an overall accuracy of 88.8%, and an AUC_{PR} of 0.68. With an average κ of 0.56 ± 0.17, an average accuracy of $88.7 \pm 4.2\%$, and an average AUC_{PR} of 0.69 ± 0.18, this scheme significantly outperforms all others, tested with a Wilcoxon test ($p < 0.0001$). The table indicates that smoothing the features per recording resulted in a significant increase in both κ and AUC_{PR} regardless of whether time delay was considered. The classification performances of the four schemes are also compared by PR curves in Fig. 5. Taking a recording as an example, Fig. 6 visually compares the PSG-based annotations and the identified classes, suggesting an enhancement in classification performance when applying feature smoothing and time delay. The figure also illustrates that feature smoothing can help removing spurious (very few epochs) detections of a class in the middle of a longer period of the other class. It confirms our expectation that the feature smoothing is an adequate way to handle this type of errors.

Table IV presents the confusion matrix of our SWS and non-SWS classifier based on cardiorespiratory features with smoothing and a 5-min negative delay. To analyze the source of false positives or alarms (i.e., instances where *non-SWS* epochs were classified as *SWS*), the breakdowns of classification results for *non-SWS* between wake, REM sleep, S1, and S2 are also given.

When using one signal modality alone, the classification performance would degrade as shown in Table V (average of $\kappa = 0.54$ for ECG or of $\kappa = 0.51$ for RE). Since the optimal time delay (-5 min) was found to be the same for either ECG

TABLE III
SUMMARY OF SWS AND NON-SWS CLASSIFICATION RESULTS IN DIFFERENT SCHEMES (TENFOLD CV)

Scheme	Result	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Kappa κ	AUC_{PR}
A: without smoothing and without time delay	Pool	53.8	53.9	91.8	86.1	0.45	0.54
	Average	53.5 ± 17.7	54.9 ± 18.7	91.8 ± 3.5	86.0 ± 4.3	0.43 ± 0.17	0.55 ± 0.18
B: with smoothing and without time delay	Pool	56.8	57.2	92.3	87.0	0.49	0.60
	Average	56.8 ± 18.3	58.1 ± 18.7	92.4 ± 3.8	87.0 ± 4.3	0.48 ± 0.17	0.61 ± 0.18
C: without smoothing and with time delay ($\tau = -2.5$ min)	Pool	59.1	61.7	92.5	87.9	0.53	0.62
	Average	59.0 ± 17.7	63.2 ± 20.5	92.5 ± 3.5	87.8 ± 4.4	0.52 ± 0.18	0.63 ± 0.18
D: with smoothing and with time delay ($\tau = -5$ min)	Pool	61.8	65.6	92.9	88.8	0.57	0.68
	Average	62.0 ± 17.8	67.2 ± 20.4	93.0 ± 3.7	88.7 ± 4.2	0.56 ± 0.17	0.69 ± 0.18

Note: Six cardiorespiratory features selected via CFS (see Table II) were used. Classifier threshold was chosen to maximize κ for training data. Significance was confirmed between scheme D and the others for accuracy, κ, and AUC_{PR} using a paired (two-sided) Wilcoxon signed-rank test ($p < 0.0001$).

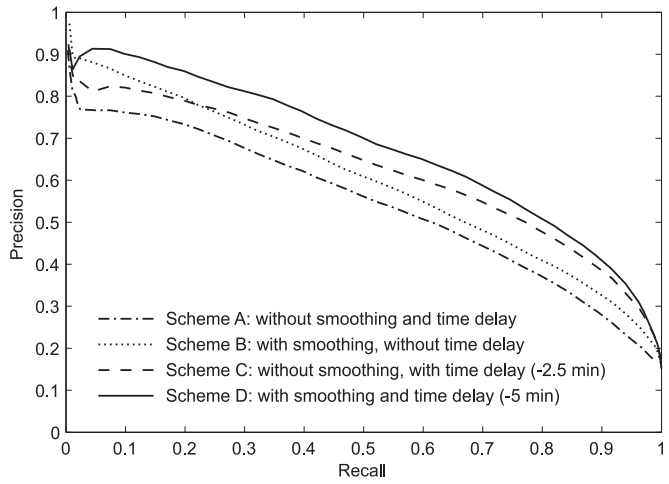


Fig. 5. Pooled PR curves of SWS and non-SWS classification using the six cardiorespiratory features selected via CFS (see Table II) and tenfold CV in different schemes. A and B compare the effect of smoothing without any time delay. C and D are the optimal delayed schemes without and with smoothing, respectively. Scheme D performed the best.

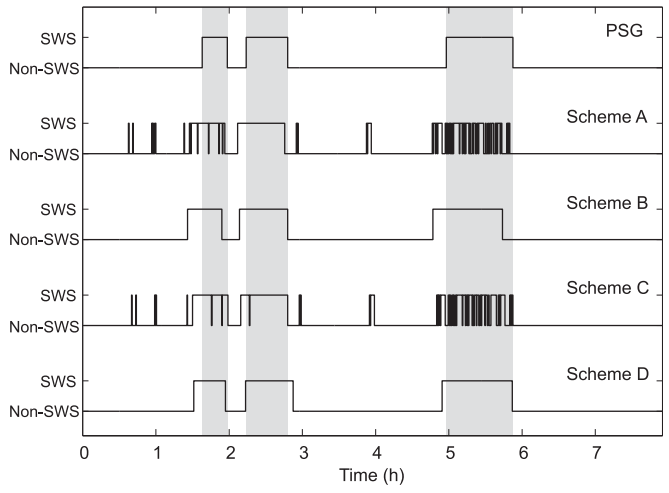


Fig. 6. Example of overnight PSG-based annotations and the corresponding SWS and non-SWS classification results using the six cardiorespiratory features selected via CFS (see Table II) in different schemes.

TABLE IV
CONFUSION MATRIX OF SWS AND NON-SWS CLASSIFICATION (TENFOLD CV) WITH INDICATION OF FALSE POSITIVES

PSG →	*SWS*	*non-SWS*				
Classified ↓		Total	S2	S1	REM	Wake
SWS	23122	14283	12841	444	424	574
non-SWS	12106	185309	90693	18172	36844	39060

Note: Six cardiorespiratory features selected via CFS (see Table II) were used.

TABLE V
COMPARISON OF SWS AND NON-SWS CLASSIFICATION RESULTS (TENFOLD CV) USING DIFFERENT SIGNAL MODALITIES

Modality	# Features*	Accuracy	Kappa κ	AUC_{PR}
ECG	4	88.2%	0.54	0.65
RE	4	87.7%	0.51	0.61

Note: Pooled results (in scheme D) are presented.
*Features were selected via CFS (see Table II).

or RE features, it was then used for comparison. Although the inclusion of ECG and RE signals yielded a better classification performance and they can be easily and unobtrusively acquired as mentioned before, our approach is still applicable to achieve reasonable results when one of them is absent.

We also applied our SWS detection approach for all the 325 recordings and for those in the low-SWS group (68 recordings from 51 subjects) with the total SWS time of less than 30 min, where the classification results are presented in Table VI. The results for the low-SWS group ($\kappa = 0.21$) were much worse than the normal group engaged in this study, due to which the classification performance for all recordings dropped to $\kappa = 0.51$. Fig. 7 (upper graph) illustrates the relation between the amount of SWS and age, confirming what is known from the literature [39], i.e., that the amount of SWS decreases with age. Fig. 7 (middle graph) illustrates the classification performance versus SWS time, which were (positively) significantly correlated. Fig. 7 (lower graph) shows a significant (negative) correlation between κ and age, indicating that the classification performance was age dependent.

TABLE VI
PERFORMANCE COMPARISON OF SWS DETECTION (TENFOLD CV) IN DIFFERENT GROUPS OF RECORDINGS IN TERMS OF OVERNIGHT SWS TIME

Group	N	SWS time	Accuracy	Kappa κ	AUC_{PR}
Normal*	257	$\geq$30 min	88.8%	0.57	0.68
Low-SWS	68	(0, 30) min	92.3%	0.21	0.17
All	325	$>$0 min	88.9%	0.51	0.58

Note: Pooled results (in scheme D) are presented. Six cardiorespiratory features selected via CFS (see Table II) were used. *The group focused in this work, in which the recordings had a more representative normal sleeping pattern as in a home-based environment where the overnight SWS time was less influenced by laboratory effects.

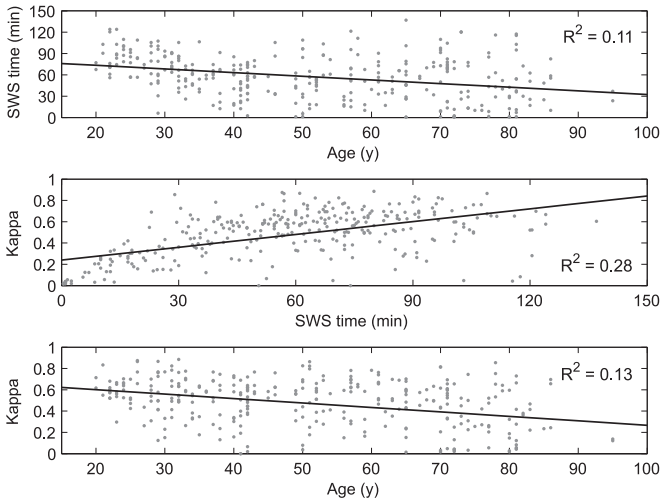


Fig. 7. Relation between overnight SWS time, subject age, and classification performance (κ) of all 325 recordings (including the normal and the low-SWS recordings). Lines represent the linear equations fitted for data from samples. Significant Spearman's rank correlation was found between SWS time and age ($r = -0.35$), between κ and SWS time ($r = 0.52$), and between κ and age ($r = -0.32$) at $p < 0.001$.

VI. DISCUSSION

It is noted that Hedner *et al.* [10] evaluated a sleep staging system and obtained a SWS and non-SWS classification with a κ of 0.48 (recomputed in terms of their reported confusion matrix). They deployed pulse rate, peripheral arterial tone, and actigraphy and their κ value is smaller than that produced in this study. To provide a fair comparison, we achieved a κ of 0.51 for all 325 recordings, which still outperforms their result. A respiratory-based sleep stager was developed in our previous work [25], reporting a κ of 0.43 in detecting SWS, where a subset of 48 normal sleep nights from the same database were included. It is lower than the result presented here ($\kappa = 0.51$), generated using only four respiratory features.

The usefulness of smoothing features with spline fitting might be similar to that of increasing the window size when computing feature values. Here, the reason of performing spline smoothing at the "feature level" was that it can interpolate missing feature values, especially for some cardiac features with missing R peaks that could not be identified due to motion artifacts. In spite of that, effects of these two smoothing approaches on SWS detection should be further analyzed.

Although feature smoothing can, in general, improve SWS detection, it may also introduce errors when detecting very short SWS periods. This is because some of the high-frequency feature components are likely not capturing noise or outliers caused by motion artifacts, but rather reflect some essential characteristics of short SWS periods. In the case of a longer SWS duration, smoothing the feature values would reduce noise, and in consequence, increase specificity. However, in the case of a shorter SWS duration (i.e., fragmented SWS), sensitivity would decrease. This procedure should then be adopted by finding an optimal tradeoff between rejecting noise and keeping useful information. As shown in Table III, the spline fitting increased all metrics. This was also the case for the low-SWS recordings where we found that the improvement was mainly contributed by the increase of specificity. The reason might be that many false positives (misclassified *non-SWS* epochs) in a long period of non-SWS were corrected through feature smoothing.

In addition, the "optimal" time delay might be varying over the night possibly influenced by the sleep stage immediately before or after SWS periods where the transition dynamics between SWS and different stages in non-SWS usually change over time [59]. Hence, fixing the time delay on features might not be the most appropriate strategy. An adaptive time-delay model depending on time of night (since lights OFF) should instead be investigated.

From Table IV, we notice that the false alarms mostly occurred in stage S2, of which 12.4% were misclassified as *SWS*. Evidence has shown that the autonomic activity differs little between S2 sleep and SWS [20]. Consequently, regardless of whether cardiac or respiratory activity was used, there might be small differences between these two sleep stages. In fact, even with PSG, manually scoring SWS (mainly based on EEG delta waves) is somewhat difficult due to the gradual changes in physiology between SWS and light sleep [35], [60]. A relatively low interrater agreement with a Cohen's Kappa coefficient of only 0.71 was reported using PSG scoring [60], which would lead to presence of fragmented SWS shown in hypnograms (PSG-based annotations). It is also suggested in that study that using the new AASM guidelines slightly higher interrater reliability for scoring SWS compared to the R&K rules can be obtained. Nevertheless, it merits further investigation upon how to better discriminate between S2 and SWS stages by means of cardiorespiratory information.

As presented in Table VI, the marked decreased classification results for the low-SWS group indicate that our classifier could not well handle any recordings with a very low number of *SWS* epochs. Fig. 7 shows that the total SWS time over night was significantly correlated with subject age (negatively) and the classification performance (positively). The figure also illustrates that our SWS detector performed better for young subjects with more SWS time than elderly. The recordings in the low-SWS group had heavily imbalanced classes, clearly more challenging from a classification point of view. Moreover, the recordings with a decreased SWS time, likely caused by the "laboratory

effect" in a laboratory study [40], might also encounter fragmented SWS. In this case, it is speculated that the autonomic changes (expressed by the cardiorespiratory features) could not be clearly observed during the "fast" and frequent changes between SWS and non-SWS, whereas they can be observed in EEG more easily. This would result in difficulty in correctly detecting *SWS* epochs from nocturnal sleep. On the other hand, the (Z-score) feature normalization assumed that all the recordings have similar percentages of sleep stages (in particular SWS in this study), which was apparently not the case for the low-SWS nights. Normalizing features for those nights would possibly lead to many false positives where many S2 sleep epochs would be wrongly classified as *SWS* epochs. In practice, the low-SWS nights should not be neglected and it is, therefore, suggested to address SWS and non-SWS classification for subjects with less overnight SWS time (likely at older ages) in future work.

Table V shows that the classification performance had relatively small drops when using only one signal modality ($\kappa = 0.54$ for ECG, $\kappa = 0.51$ for RE) compared with using both modalities. However, although the use of both cardiac and respiratory signals yields a better classification performance and they can be easily and unobtrusively acquired, our approach is still applicable to achieve a reasonable classification performance when one of them is absent. This is particularly important from the perspective of unobtrusive SWS detection so that an unobtrusive sensing system can be selectively designed to acquire only one of these two signal modalities or both of them upon the request of classification performance.

As stated, this study focused on healthy adults without sleep disorders so that the methods and the corresponding settings (e.g., parameters when computing features, selected features, the use of normalization, smoothing window size, and time delay) were optimally tuned on the data from healthy subjects and might not be appropriate for the other subject groups with prevalent sleep problems such as sleep apnea. It is known that patients with the sleep apnea syndrome are associated with altered cardiac variability [61], decreased SWS [30], sleep fragmentation (in particular fragmented SWS) due to the repeated occurrences of end-apneic arousals [62], altered sleep stage transition dynamics [63], and dysfunction in autonomic nervous activity [64]. Moreover, healthy adults can also have occasional arrhythmia during sleep, e.g., related to stress, hormonal, metabolic, and autonomic events, which were not been taken into account in this study. All these would, as a consequence, be barriers in cardiorespiratory-based SWS detection, which requires further investigation.

In practice, automatic cardiorespiratory-based SWS detection can potentially enable the prediction of SWS onset in an early manner using unobtrusive sensors as a form of, for example, a wrist-worn watch/band or a textile bedsheet as mentioned before. It will benefit the applications of prompting slow waves for enhancing memory consolidation during sleep in a home scenario, which requires a system for online SWS detection. Our proposed approach can anticipate the occurrence of SWS with 5 min ahead where the window size (one side) used for computing features were all less than this time interval. Although the feature normalization (based on the entire-night recording) and smoothing (with a spline window of 25 epochs) would still limit the achievement of an online SWS detector, these limitations seem manageable. First, since the feature normalization mainly served to diminish between-subject variability, this can be alternatively achieved by using the previous night (as a baseline) to normalize the following nights in real applications where usually multiple nights are expected. Second, the smoothing window size can be reduced and the smoothing can be "time progressive" to fit the online requirement. Nevertheless, their influences on the detection performance should be further studied when targeting an online SWS detection system.

VII. Conclusion

In this study, overnight epoch-by-epoch classification of nocturnal SWS and non-SWS was achieved based on cardiorespiratory signals that can be acquired unobtrusively. To reduce classification errors caused by, for example, sensor noise, body motion artifacts, and/or within-subject variability, a recording-specific feature smoothing using spline fitting was employed. Besides, we used the features anticipating each target epoch to identify SWS of that epoch as long as the preceding cardiorespiratory activity (compared with the PSG-based annotations) appeared during the transitions between SWS and non-SWS. With an LD classifier, we revealed that the use of feature smoothing and time delay profoundly improved the classification performance (κ of 0.57 versus 0.45). Our approach also produced reasonable classification results when only one of the signal modalities was present. Furthermore, the classifier performed much better for subjects who had more total SWS time than for subjects with less SWS time.

References

[1] E. A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, National Institutes of Health, Washington, DC, USA, 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 and Technical Specifications*, American Academy of Sleep Medicine, Westchester, IL, USA, 2007.

[3] R. J. Berger and N. H. Phillips, "Energy conservation and sleep," *Behav. Brain Res.*, vol. 69, no. 1–2, pp. 65–73, Aug. 1995.

[4] R. Stickgold, "Sleep-dependent memory consolidation," *Nature*, vol. 437, pp. 1272–1278, Oct. 2005.

[5] M. H. Bonnet, "Effect of sleep disruption on sleep, performance, and mood," *Sleep*, vol. 8, no. 1, pp. 11–19, 1985.

[6] E. Tasali, R. Leproult, D. A. Ehrmann, and E. V. Cauter, "Slow-wave sleep and the risk of type 2 diabetes in humans," in *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, no. 3, pp. 1044–1049, Jan. 2008.

[7] L. Marshall, H. Helgadóttir, M. Mölle, and J. Born, "Boosting slow oscillations during sleep potentiates memory," *Nature*, vol. 444, pp. 610–613, Nov. 2006.

[8] M. Massimini *et al.*, "Triggering sleep slow waves by transcranial magnetic stimulation," in *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, no. 20, pp. 8496–8501, Mar. 2007.

[9] H.-V. V. Ngo, T. Martinetz, J. Born, and M. Mölle, "Auditory closed-loop stimulation of the sleep slow oscillation enhances memory," *Neuron*, vol. 78, no. 3, pp. 545–553, May 2013.

[10] J. Hedner *et al.*, "Sleep staging based on autonomic signals: a multi-center validation study," *J. Clin. Sleep Med.*, vol. 7, no. 3, pp. 310–316, Jun. 2011.

[11] J. M. Kortelainen, M. O. Mendez, A. M. Bianchi, M. Matteucci, and S. Cerutti, "Sleep staging based on signals acquired through bed sensor," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 776–785, May 2010.

[12] L. Samy, M.-C. Huang, 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, Jul. 2014.

[13] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[14] D. Pevernagie, R. M. Aarts, and M. D. Meyer, "The acoustics of snoring," *Sleep Med. Rev.*, vol. 14, no. 2, pp. 131–144, Apr. 2010.

[15] 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, no. 3, pp. 627–634, Mar. 2012.

[16] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Feb. 2007.

[17] M. H. Bonnet and D. L. Arand, "Heart rate variability: Sleep stage, time of night, and arousal influences," *Electroencephalogr. Clin. Neurophysiol.*, vol. 102, no. 5, pp. 390–396, May 1997.

[18] 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, Nov. 1982.

[19] A. Baharav *et al.*, "Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability," *Neurology*, vol. 45, no. 6, pp. 1183–1187, Jun. 1995.

[20] J. Trinder *et al.*, "Autonomic activity during human sleep as a function of time and sleep stage," *J. Sleep Res.*, vol. 10, no. 4, pp. 253–264, Dec. 2001.

[21] P. Bušek, J. Vaňková, J. Opavský, J. Salinger, and S. Nevšímalová, "Spectral analysis of the heart rate variability in sleep," *Physiol. Res.*, vol. 54, pp. 369–376, 2005.

[22] S. J. Redmond, P. de Chazal, C. O'Brien, S. Ryan, W. T. McNicholas, and C. Heneghan, "Sleep staging using cardiorespiratory signals," *Somnologie*, vol. 11, pp. 245–256, Dec. 2007.

[23] W. Karlen, C. Mattiussi, and D. Floreano, "Sleep and wake classification with ECG and respiratory effort signals," *IEEE Trans. Biomed. Circuits Syst.*, vol. 3, no. 2, pp. 71–78, Apr. 2009.

[24] M. O. Mendez, M. Matteucci, V. Castronovo, L. Ferini-Strambi, S. Cerutti, and A. M. Bianchi, "Sleep staging from heart rate variability: time-varying spectral features and hidden Markov models," *Int. J. Biomed. Eng. Technol.*, vol. 3, no. 3/4, pp. 246–263, 2010.

[25] 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, Nov. 2014.

[26] T. Willemen *et al.*, "An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 2, pp. 2168–2194, 2014.

[27] S. Herscovici, A. Pe'er, S. Papyan, and P. Lavie, "Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy," *Physiol. Meas.*, vol. 28, no. 2, pp. 129–140, Feb. 2007.

[28] M. Bresler, K. Sheffy, G. Pillar, M. Preiszler, and S. Herscovici, "Differentiating between light and deep sleep stages using an ambulatory device based on peripheral arterial tonometry," *Physiol. Meas.*, vol. 29, no. 5, pp. 571–584, May 2008.

[29] Z. Shinar, A. Baharav, Y. Dagan, and S. Akselrod, "Automatic detection of slow-wave-sleep using heart rate variability," in *Proc. Comput. Cardiol.*, Rotterdam, The Netherlands, Sep. 2001, pp. 593–596.

[30] M. A. Carskadon and W. C. Dement, "Monitoring and staging human sleep—Chapter 2: Normal human sleep," in *Principles and Practice of Sleep Medicine*, 5th ed., M. H. Kryger, T. Roth, and W. C. Dement, Eds. St. Louis, MO, USA: Elsevier Saunders, 2011, ch. 2, pp. 16–26.

[31] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Proc. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[32] C. De Boor, *A Practical Guide to Splines*. New York, NY, USA: Springer, 2001.

[33] H. Otzenberger, C. Simon, C. Gronfier, and G. Brandenberger, "Temporal relationship between dynamic heart rate variability and electroencephalographic activity during sleep in man," *Neurosci. Lett.*, vol. 229, no. 3, pp. 173–176, Jul. 1997.

[34] F. Jurysta *et al.*, "A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men," *Clin. Neurophysiol.*, vol. 114, no. 11, pp. 2146–2155, Nov. 2003.

[35] X. Long, J. B. Arends, R. M. Aarts, R. Haakma, P. Fonseca, and J. Rolink, "Time delay between cardiac and brain activity during sleep transitions," *Appl. Phys. Lett.*, vol. 106, no. 14, pp. 143 702-1–143 702-4, Apr. 2015.

[36] 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, Jul. 2014.

[37] X. Long, P. Fonseca, R. Haakma, J. Foussier, and R. M. Aarts, "Automatic detection of overnight deep sleep based on heart rate variability: A preliminary study," in *Proc. IEEE 36th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Chicago, IL, USA, Aug. 2014, pp. 50–53.

[38] G. Klösch *et al.*, "The siesta project polygraphic and clinical database," *IEEE Signal Proc. Mag.*, vol. 20, no. 3, pp. 51–57, May/Jun. 2001.

[39] M. M. Ohayon, M. A. Carskadon, C. Guilleminault, and M. V. Vitiello, "Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan," *Sleep*, vol. 27, no. 7, pp. 1255–1273, Nov. 2004.

[40] J. Mendels and D. R. Hawkins, "Sleep laboratory adaptation in normal subjects and depressed patients ("first night effect")," *Electroencephalogr. Clin. Neurophysiol.*, vol. 22, no. 6, pp. 556–558, Jun. 1967.

[41] 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, May 2006.

[42] J. A. van Alsté, W. van Eck, and O. E. Herrmann, "ECG baseline wander reduction using linear phase filters," *Comput. Biomed. Res.*, vol. 19, pp. 417–427, Oct. 1986.

[43] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 12, pp. 1157–1165, Dec. 1986.

[44] P. Fonseca, R. M. Aarts, J. Foussier, and X. Long, "A novel low-complexity post-processing algorithm for precise QRS localization," *SpringerPlus*, vol. 3, art. no. 376 (13 pages), Jul. 2014.

[45] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.

[46] A. M. Bianchi, L. Mainardi, E. Petrucci, M. G. Signorini, M. Mainardi, and S. Cerutti, "Time-variant power spectrum analysis for the detection of transient episodes in HRV signal," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 2, pp. 136–144, Feb. 1993.

[47] A. Boardman, F. S. Schlindwein, A. P. Rocha, and A. Leite, "A study on the optimum order of autoregressive models for heart rate variability," *Physiol. Meas.*, vol. 23, no. 2, pp. 325–336, Mar. 2002.

[48] X. Long, P. Fonseca, R. Haakma, R. M. Aarts, and J. Foussier, "Time-frequency analysis of heart rate variability for sleep and wake classification," in *Proc. IEEE 12nd Int. Conf. Bioinform. Bioeng.*, Larnaca, Cyprus, Nov. 2012, pp. 85–90.

[49] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis," *Physica A, Stat. Mech. Appl.*, vol. 295, no. 3–4, pp. 441–454, Jun. 2001.

[50] T. Penzel, J. W. Kantelhardt, L. Grote, J. H. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1143–1151, Oct. 2003.

[51] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E*, vol. 71, no. 2, pp. 021 906-1–021 906-18, Feb. 2005.

[52] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol. Heart Circ. Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.

[53] X. Long *et al.*, "Measuring dissimilarity between respiratory effort signals based on uniform scaling for sleep staging," *Physiol. Meas.*, vol. 35, no. 12, pp. 2529–2542, Dec. 2014.

[54] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, Hamilton, New Zealand, 1999.

[55] X. D. Zhang, "Strictly standardized mean difference, standardized mean difference and classical t-test for the comparison of two groups," *Stat. Biopharm. Res.*, vol. 2, no. 2, pp. 292–299, 2010.

[56] P. C. Mahalanobis, "On the generalised distance in statistics," *Proc. Natl. Inst. Sci. India*, vol. 2, no. 1, pp. 49–55, 1936.

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

[58] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, vol. 10, no. 2, 2006, pp. 233–240.
[59] A. Kishi, Z. R. Struzik, B. H. Natelson, F. Togo, and Y. Yamamoto, "Dynamics of sleep stage transitions in healthy humans and patients with chronic fatigue syndrome," *Amer. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 294, no. 6, pp. R1980–R1987, Jun. 2008.
[60] H. Danker-Hopfe *et al.*, "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard," *J. Sleep Res.*, vol. 18, no. 1, pp. 74–84, Mar. 2009.
[61] K. Narkiewicz, N. Montano, C. Cogliati, P. J. van de Borne, M. E. Dyken, and V. K. Somers, "Altered cardiovascular variability in obstructive sleep apnea," *Circulation*, vol. 98, no. 11, pp. 1071–1077, Sep. 1998.
[62] R. J. Kimoff, "Sleep fragmentation in obstructive sleep apnea," *Sleep*, vol. 19, no. 9 Suppl, pp. S61–S66, Nov. 1996.
[63] M. T. Bianchi, S. S. Cash, J. Mietus, C.-K. Peng, and R. Thomas, "Obstructive sleep apnea alters sleep stage transition dynamics," *PLOS One*, vol. 5, no. 6, pp. e11 356-1–e11 356-12, Jun. 2010.
[64] C. Gulleminault, A. T. Lehrman, L. Forno, and W. C. Dement, "Sleep apnoea syndrome: States of sleep and autonomic dysfunction," *J. Neurol. Neurosurg. Psychiatry*, vol. 40, no. 7, pp. 718–725, Jul. 1977.

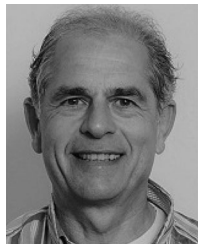
Xi Long (M'09) was born in Ganzhou, China, in 1983. He received the B.Eng. degree in electronic information engineering from Zhejiang University, Hangzhou, China, in 2006 and the M.Sc. and the Ph.D. (cum laude) degrees in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2009 and 2015, respectively.

He has five years of R&D experience in healthcare industry, working for Philips Research, Eindhoven. He is currently a Research Scientist at Philips Research and the Eindhoven University of Technology. His research interests include unobtrusive/wearable sensing, medical signal analysis, and machine learning in health and medicine.

Dr. Long received the Best Student Paper Award from the IEEE Conference on Bioinformatics and Bioengineering and the First Runner-up Paper Award from the IEEE Conference on Biomedical and Health Informatics in 2012.

Pedro Fonseca was born in Lisbon, Portugal, in 1979. He received the Dipl.-Ing. degree in electrical engineering and the M.Sc. degree in electrical and computer engineering from the Technical University of Lisbon, Lisbon, Portugal.

He has ten years of R&D experience at Philips Research, Eindhoven, The Netherlands, in the fields of image and video analysis and, more recently, personal healthcare. His current research interests include biosignal processing, machine learning, and unobtrusive physiological sensing.



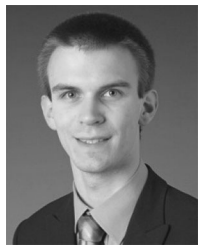
Ronald M. Aarts (M'95–SM'95–F'07) was born in Amsterdam, The Netherlands, in 1956. He received the B.Sc. degree in electrical engineering and the Ph.D. degree in physics from the Delft University of Technology, Delft, The Netherlands, in 1977 and 1995, respectively.

He joined the Optics group, Philips Research Laboratories, Eindhoven, in 1977 and initially investigated servos and signal processing for use in both Video Long Play players and Compact Disc players. In 1984, he joined the Acoustics group at Philips Research Labs and worked on the development of CAD tools and signal processing for loudspeaker systems. In 1994, he became a member of the Digital Signal Processing (DSP) group at Philips Research and has led research projects on the improvement of sound reproduction, by exploiting DSP and psycho-acoustical phenomena. In 2003, he became a Philips Fellow at Philips Research, and extended his interests in engineering to medicine and biology in particular sensors, signal processing, and systems for unobtrusive monitoring, sleep, and epilepsy detection. Since 2006, he is a part-time Full-Professor at the Eindhoven University of Technology, Eindhoven. He has published more than three hundred papers and reports, and holds more than 190 first patent application filings including more than 60 granted US-patents in the aforementioned fields.

Dr. Aarts received the Gilles Hold Award in 1999, the Gold Invention Award in 2012, and the Eureka Award in 2001 and 2014 from Philips. He has served on a number of organizing committees and as a Chairman for various international conventions. He received Silver Medal from Audio Engineering Society , and is a Fellow and past Governor. He is a member of the NAG (Dutch Acoustical Society), the Acoustical Society of America, the Dutch Society for Biophysics and Biomedical Engineering, and the NSWO (Dutch Society for Sleep and Wake Research).

Reinder Haakma received the M.Sc. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 1985, and the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1998.

He is currently a Principal Scientist with Philips Research, Eindhoven. His research interests include unobtrusive monitoring of health and sleep.



Jérôme Rolink (M'11) was born in Cologne, Germany, in 1984. He received the Dipl.-Ing. degree in electrical engineering from RWTH Aachen University, Aachen, Germany, where he is currently working toward the Dr.-Ing. (Ph.D.) degree at the Philips Chair of Medical Information Technology.

He is currently a Research Assistant at RWTH Aachen University. His research interests include signal processing and classification as well as physiological measurement techniques.

Steffen Leonhardt (M'95–SM'06) was born in Frankfurt, Germany, in 1961. He received the M.S. degree in computer engineering from the State University of New York, Buffalo, NY, USA in 1987, the Dipl.Ing. degree in electrical engineering, in 1989 and the Dr. Ing. degree in control engineering from the Technical University of Darmstadt, Darmstadt, Germany, in 1995, and the M.D. degree in medicine from J. W. Goethe University, Frankfurt, Germany, in 2001.

He has five years of R&D management experience in medical engineering industry working for Dräger Medical AG & Co KGaA, Lübeck, Germany. In 2003, he was appointed as a Full Professor and the Head of the Philips endowed Chair of Medical Information Technology at RWTH Aachen University, Aachen, Germany.

Dr. Leonhardt serves as an Associate Editor of the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS and IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS. In 2014, he became a Fellow of the NRW Academy of Sciences, Humanities and the Arts, Düsseldorf, Germany. In 2015, he became the Distinguished Lecturer with the IEEE Engineering in Medicine and Biology Society.