



Citation/Reference	<p>Carolina Varon, Alexander Caicedo, Dries Testelmans, Bertien Buyse, Sabine Van Huffel (2015),</p> <p>A novel algorithm for the automatic detection of sleep apnea from single-lead ECG</p> <p>IEEE Transactions on Biomedical Engineering (TBME), Accepted.</p>
Archived version	<p>Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher</p>
Published version	<p>Not yet available</p>
Journal homepage	<p>http://tbme.embs.org/</p>
Author contact	<p>Carolina.varon@esat.kuleuven.be</p> <p>your phone number + 32 (0)16 32 64 17</p>
IR	<p>Not yet available</p>

(article begins on next page)



A novel algorithm for the automatic detection of sleep apnea from single-lead ECG

Carolina Varon, *Member, IEEE*, Alexander Caicedo, *Member, IEEE*, Dries Testelmans, Bertien Buyse, Sabine Van Huffel, *Fellow, IEEE*

Abstract—Goal: This paper presents a methodology for the automatic detection of sleep apnea from single-lead ECG. **Methods:** It uses two novel features derived from the ECG, and two well-known features in heart rate variability analysis, namely the standard deviation and the serial correlation coefficients of the RR interval time series. The first novel feature uses the principal components of the QRS complexes, and it describes changes in their morphology caused by an increased sympathetic activity during apnea. The second novel feature extracts the information shared between respiration and heart rate using orthogonal subspace projections. Respiratory information is derived from the ECG by means of three state-of-the-art algorithms, which are implemented and compared here. All features are used as input to a least-squares support vector machines (LS-SVM) classifier, using an RBF kernel. In total, 80 ECG recordings were included in the study. **Results:** Accuracies of about 85% are achieved on a minute-by-minute basis, for two independent datasets including both hypopneas and apneas together. Separation between apnea and normal recordings is achieved with 100% accuracy. In addition to apnea classification, the proposed methodology determines the contamination level of each ECG minute. **Conclusion:** The performances achieved are comparable with those reported in the literature for fully automated algorithms. **Significance:** These results indicate that the use of only ECG sensors can achieve good accuracies in the detection of sleep apnea. Moreover, the contamination level of each ECG segment can be used to automatically detect artefacts, and to highlight segments that require further visual inspection.

Index Terms—Sleep Apnea, ECG Morphology, Cardiorespiratory interactions, LS-SVM

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC); PhD/Postdoc grants Flemish Government: FWO: projects: G.0427.10N (Integrated EEG-fMRI), G.0108.11 (Compressed Sensing) G.0869.12N (Tumor imaging) G.0A5513N (Deep brain stimulation); PhD/Postdoc grants IWT: projects: TBM 080658-MRI (EEG-fMRI), TBM 110697-NeoGuard; PhD/Postdoc grants. iMinds Medical Information Technologies SBO 2014, ICON: NXT_Sleep. Flanders Care: Demonstratieproject Tele-Rehab III (2012-2014). Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, “Dynamical systems, control and optimization”, 2012-2017). Belgian Foreign Affairs-Development Cooperation: VLIR UOS programs. EU: The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n°339804). This paper reflects only the authors’ views and the Union is not liable for any use that may be made of the contained information. Other EU funding: RECAP 209G within INTERREG IVB NWE programme, EU MC ITN TRANSACT 2012 (n°316679), ERASMUS EQR: Community service engineer (n°539642-LLP-1-2013)

C. Varon, A. Caicedo, and S. Van Huffel are with the Department of Electrical Engineering-ESAT, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, and iMinds Medical IT Department, KU Leuven, B-3001 Leuven, Belgium, (e-mail: carolina.varon@esat.kuleuven.be)

D. Testelmans and B. Buyse are with the UZ Leuven, Department of Pneumology, Leuven, Belgium.

I. INTRODUCTION

SLEEP apnea is a sleep-related breathing disorder that affects about 10% of middle aged adults [1]. In addition, it is considered a risk factor for morbidity and mortality due to its long-term effect on the cardiovascular system [2]. This effect is related to different physiological mechanisms like systemic hypertension and increased sympathetic modulation that in a long term compromises the well-functioning of the heart.

Apnea and hypopnea are two types of respiratory events characterized by a complete absence or reduction of airflow, respectively, during at least 10s [3]. During these events the respiratory effort can differ, and this on its turn can be used to distinguish between different types of apnea. For instance, in obstructive sleep apnea (OSA) the patients stop breathing, despite the respiratory effort, due to an obstruction in the upper airways that avoids air entering the lungs. In central sleep apnea (CEN), patients also stop breathing, but in this case due to absence of respiratory effort. The combination of these two apnea types is called mixed apnea (MIX), and it consists of a reduction of respiratory effort that leads to an upper airway obstruction. Hypopneas are characterized by a drop in oxygen saturation caused by a reduction in the airflow. In addition, no clear ECG changes can be observed in the majority of the cases. As a result, hypopneas are the most challenging type of respiratory events to be detected solely from the ECG signal. In these cases other signals like oxygen saturation and airflow can complement the ECG analysis. Apneas and hypopneas can occur up to hundreds of times per night, and its recurrence can cause hypersomnolence, excessive daytime sleepiness, insomnia, nocturia, memory loss, attention deficit, and depression among others.

Currently, sleep-related breathing disorders are diagnosed using poly(somno)graphy, which is a sleep test performed in a hospital or at home, and where the evaluation and supervision of a clinician is essential. Even though poly(somno)graphy is the most important diagnostic tool in sleep medicine, its high costs and low comfort significantly reduce its diagnostic power. Therefore, the development of more portable and less intrusive systems is of paramount importance.

During poly(somno)graphy several physiological parameters are recorded such as heart rate and respiration, which are deeply affected during episodes of apnea. Hence, several studies have focused on the detection of sleep apnea using as few signals as possible. For example, [4],[5],[6] and [7] proposed to derive several features from the electrocardiographic (ECG) signal to detect sleep apnea. Moreover, instead

of measuring the respiratory signal, those algorithms first estimate the respiration from the ECG, and then derive a set of features that can later be used for the discrimination of apneas.

In this context, this study proposes a simple algorithm for the automatic detection of sleep apnea from the ECG signal. In addition, two different features that capture valuable information related to apneas are proposed. One describes changes in the morphology of the ECG, and one computes the information shared between respiration and heart rate, by means of orthogonal subspace projections. Changes in the morphology of the ECG associated with obstructive sleep apnea events were first reported in [8], where it was concluded that algorithms that take into account changes in morphology perform better for the classification of apnea. Later in [9], changes in the P-waves were used to discriminate between normal ECG segments, and segments containing apneas. In this work, however, a novel methodology based on principal component analysis applied to the QRS complexes is proposed. Apart from the two novel features, this study compares three different algorithms to compute the ECG derived respiration (EDR). The first EDR algorithm is based on the R-peak amplitude [6], which has been widely used for apnea detection, the second one uses principal component analysis [10], and the last one uses its kernel version kPCA [11]. The last two algorithms have not been applied yet to the detection of sleep apnea, and this study aims to determine their added value in this type of applications. All the features derived from the heart rate and from the EDR signals are used as input to different classifiers, namely, linear discriminant analysis (LDA), support vector machines (SVM), and least-squares support vector machines (LS-SVM). Variations of these three classifiers are implemented and compared in this work. A final algorithm is then proposed and tested on two different datasets. An important addition of the proposed methodology is that a level of confidence is provided for each classified segment. This level is associated with the level of contamination of the ECG, which is used to derive all features.

The remainder of this paper is organized as follows. Section II describes the two datasets used in this study, the pre-processing steps, the three EDR algorithms implemented, and the two novel features. The results are presented in Section III, and then discussed in section IV. Conclusions are presented in section V.

II. METHODS

A. Datasets

Two different datasets were used in this study. The first one is the Apnea-ECG database [4], [8], publicly available on Physionet [12]. This dataset contains 70 single-lead ECG signals sampled at 100Hz, of lengths ranging between 7h and 10h. Each ECG signal was manually annotated by an expert, on a minute-by-minute basis, for sleep apnea and hypoapnea events using additional signals such as respiration and oxygen saturation. Unfortunately, no differentiation between hypoapneas and apneas was indicated in the annotations. Eight of the 70 ECG recordings are accompanied by concomitant respiratory effort measured with respiratory belts in the abdomen and

TABLE I
CHARACTERISTICS OF PATIENTS IN THE LEUVEN DATASET.

Patient	Gender	BMI	AHI
1	M	40.8	72.6
2	F	26.3	3.2
3	M	21.7	0.6
4	M	30.5	42.7
5	M	27.8	45.6
6	M	29.4	26.5
7	M	31.3	29
8	M	30.9	29.6
9	M	26.6	36.2
10	M	36.6	44.2

thorax. In total, 34 324 annotated minutes were extracted and used in this study.

The second dataset was recorded in the Sleep Laboratory of the University Hospital Leuven (UZ Leuven) in Belgium, and consists of 10 single-lead ECG (lead II) signals extracted from polysomnographic recordings of 10 different patients. The mean age of the patients was 48.4 ± 11.2 , and they presented habitual snoring, witnessed apneas and hypersomnolence. The body mass index (BMI), gender, and apnea-hypopnea index for each patient are indicated in Table I. Information about respiratory effort was acquired using two respiratory belts placed around the thorax and the abdomen. All signals were sampled at 200Hz and their duration ranges from 320 to 469 minutes (384 ± 47). In total 5956 minutes were extracted and annotated according to the *Chicago Criteria* [3], by a technician and a physician experienced on interpreting polysomnographic signals. These annotations correspond to the relative and absolute time of the event, duration, and type of respiratory event: obstructive apnea (OSA); central apnea (CEN); obstructive hypopnea (OSH); hypopnea (HPA); or mixed apnea (MIX). The events annotated as OSA and HPA are characterized by a reduced or complete absence of airflow respectively, during at least 10 seconds. As mentioned in the introduction, the detection of airflow reductions can be made from signals like the oxygen saturation and airflow. A central apnea (CEN) consists of a cessation of respiratory effort, and mixed apneas (MIX) are events with initial absence of respiratory effort that are accompanied by upper airway obstructions. A hypopnea was classified as obstructive (OSH) if there was presence of snoring during the event or increased inspiratory flattening of the nasal pressure flow signal compared to baseline breathing or associated thoracoabdominal paradox during the event but not during pre-event breathing.

B. ECG pre-processing and the construction of the tachogram

Each ECG signal is segmented into epochs of 60s. This is done based on the results reported in [13], where different epoch lengths were analyzed. It was concluded that segments of 60s allow to obtain the best performance when discriminating between respiratory events.

After segmenting the ECG, the power line interference at 50Hz is filtered out using a notch filter, and the mean of the signals is removed. Next, all ECG signals are up-sampled at 250Hz [14] by means of cubic spline interpolation.

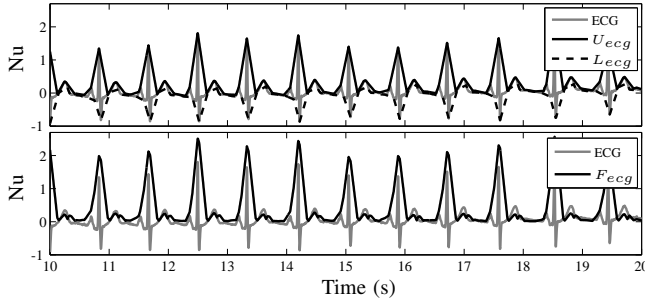


Fig. 1. Computation of the flattened ECG signal. (top) upper (U_{ecg}) and lower (L_{ecg}) envelope computed using the secant method on 150ms. (bottom) Resultant flattened ECG ($F_{ecg} = U_{ecg} - L_{ecg}$) that is used to search for the R-peaks. Remember that the locations found in F_{ecg} are used to determine the final R-peak positions in the original ECG signal. Nu stands for normalized units.

This is done in order to fall within the frequency range recommended in [14]: 250Hz-500Hz. These R-peaks are found using an adapted version of the Pan-Tompkins algorithm that is proposed in this study, where no filters (bandpass, derivative, integrator) are applied to the ECG. Instead, the upper (U_{ecg}) and lower (L_{ecg}) ECG envelopes are computed and used to obtain a flattened version defined as: $F_{ecg} = U_{ecg} - L_{ecg}$. The computation of U_{ecg} and L_{ecg} is done by means of the secant method on 150ms, and it is illustrated in Fig. 1. F_{ecg} is only used to find the possible locations of the R-peaks, which are then used to find the final R-peaks in the original ECG. In order to correct for missing, ectopic and false peaks, the search back procedure described in [6] is applied. The final location of the R-peaks is then used to compute the RR interval time series, RR , or so-called tachogram [14].

C. ECG contamination level

It is well known that different types of artefacts contaminate the ECG signals. Hence, when deriving signals like the tachogram it is crucial to determine whether an ECG segment is contaminated or not. By doing so, it is possible to estimate the reliability of the results obtained from that particular segment. To do this, the method proposed in [15] is used, where a weight ω is assigned to each ECG segment depending on how similar its autocorrelation function (ACF) is to the ACFs of the past segments. The similarity value for each segment is calculated using the cosine pair-wise similarity between ACFs, and its weight is then computed as the normalized sum of its pair-wise similarities. The closer ω is to one, the cleaner the segment is from artefacts. In [15], all segments with weights below a certain threshold (95th percentile) were removed from the dataset. Here, however, no threshold is applied. Instead, ω is used to provide a contamination level for each particular segment, and only segments with a certain value of ω are included in the training phase.

D. ECG-derived respiration

It has been shown by several authors [5]–[7], [16], [17] that respiratory events can be detected using an ECG derived respiration (EDR). The derivation of this approximated respiratory signal is possible due to the mechanical interaction

between the respiratory movements and the morphology of the ECG. This is to say that during each breathing cycle the electrical impedance of the thorax, and the relative position of the electrodes with respect to the heart change due to variations of the amount of air in the lungs. These variations can be detected from the changes in the amplitude of the different waves of the ECG.

Here, three different methodologies to compute the EDR signal are implemented, and they will be described below. The performances of the algorithm using each EDR will be assessed and compared against the one obtained using the real, reference respiratory signals. This is done in order to determine the added value of each methodology on the detection of sleep apnea from portable ECG-based monitoring systems.

1) *R-peak amplitude (Ramp)*: This methodology, first described in [18], uses the amplitude of the R-peaks in a baseline corrected ECG signal. The baseline is computed by means of two median filters, namely one of 200ms that removes the QRS complexes and P-waves, and one of 600ms that removes the T-waves. The resulting baseline is then subtracted from the original ECG, and the amplitudes of the R-peaks make up the EDR. In the remaining of this document, this EDR will be denoted by R_{amp} .

2) *Principal component analysis (PCA)*: The EDR signal computed using PCA takes into account not only the variations of the R-peak amplitude, but also the modulation of the whole QRS complex due to respiratory movements. Here, all QRS complexes are segmented using a symmetric window of 120ms around the R-peaks. The length of this window was selected based on the width of a normal QRS complex: $100ms \pm 20ms$ [19]. Next, all windows are aligned with respect to the R-peaks in a matrix X [10]. Since the amplitude of all points around the R-peaks are modulated by respiration, a mean variation is derived by means of PCA. This mean corresponds to the R_{pca} , which is the first principal component obtained from the eigendecomposition of the covariance matrix of X : $\Sigma = XX^T$. This methodology is summarized in Fig. 2.

3) *Kernel principal component analysis (kPCA)*: In [11] it was shown that a non-linear version of PCA, i.e., kPCA, achieves better approximations of the real respiratory signal, when compared with the two previous methodologies. In this case, a non-linear transformation of the input space is performed by means of an RBF kernel function, which allows to take into account linear and non-linear interactions between respiration and the morphology of the QRS complexes. The input space corresponds to the 120ms windows that were segmented around the R-peaks, and the new, transformed space is called the feature space. It is in this new space that the traditional linear PCA is performed. After that, the EDR is computed in the feature space and then mapped back to the input space by means of pre-image [20]. Details on how to compute the EDR using this methodology, namely R_{kpca} , can be found in [11].

E. Feature based on the QRS morphology

A new feature derived from the ECG is used here for the detection of sleep apnea. This feature is based on the

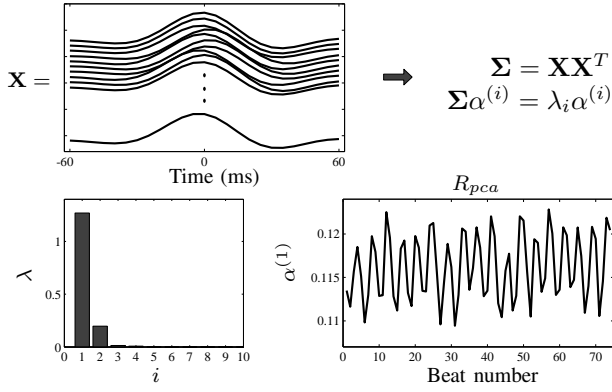


Fig. 2. Computation of the ECG-derived respiratory signal (EDR) using PCA [10]. (top) Matrix containing the QRS complexes, segmented with a window of 120ms around the R-peaks. On the right is the general procedure to compute the principal components. (bottom) The eigenvalues of the covariance matrix Σ are indicated on the left, and the eigenvector corresponding to the first eigenvalue is indicated on the right. Note that $\alpha^{(1)}$ explains most of the variance of the data, and it is identified as the EDR: R_{pca} .

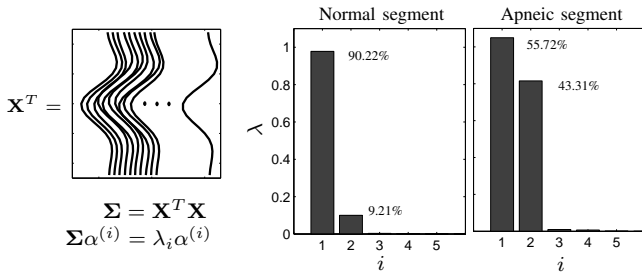


Fig. 3. Eigenvalues of the covariance matrix of X^T . The percentage of variance explained by each component is indicated next to the first two eigenvalues for both cases: apnea and normal. Note that in order to describe most of the variance contained in the morphology of the QRS complexes for an apneic segment, it is necessary to also consider the second principal component.

computation of the EDR using PCA. The variation here is that the algorithm is applied to the matrix X^T rather than to X . The reason for this relies on the fact that when considering X^T , a “local” interpretation of changes in morphology around the R-peaks can be made. Furthermore, when an episode of apnea occurs, the changes in QRS morphology within one segment are more heterogeneous, an effect that can be observed from the eigenvalues of X^T . This is illustrated in Fig. 3, where the eigenvalues obtained from two different ECG segments are indicated. The percentage of the variance explained by each component is also indicated, and it is defined as $(\lambda_i / \sum \lambda_i)$, where λ_i is the i th eigenvalue of the covariance matrix of X^T , $i = 1, \dots, m$, and m the window size. Note that during a normal episode (i.e., non apneic) the variance explained by the first component describes more than 90% of the modulation of the QRS complexes. However, when apnea occurs, it is necessary to also consider the information contained in the second component, which explains more than 40% of the variance. Following these observations, the percentage explained by the second component is computed for each ECG segment of 60s, and it is used as a “morphology” feature during the classification phase.

The procedure discussed above is strongly connected to the singular value decomposition of the matrix X , which is defined

as $X = USV^T$, where U are the eigenvectors of XX^T , V the eigenvectors of $X^T X$ and the non-zero elements of the matrix S correspond to the square root of the eigenvalues of XX^T or $X^T X$. However, the PCA approach was adopted here in order to include a direct link with the EDR methodology.

F. Feature derived from subspace projections

It is well known that variations in heart rate are modulated by variations in respiration through the physiological mechanism called respiratory sinus arrhythmia (RSA) [21]. Therefore, the information contained in the heart rate signal could in principle be split into two different components, one related to respiratory changes, and one related to physiological mechanisms different from respiration (e.g., sympathetic activation). This decomposition can be computed by means of orthogonal subspace projections, as proposed in [22][23], and it will be described as follows.

Given are two physiological signals X and Y , where X is the RR interval time series RR (resampled at 5Hz), and Y corresponds to one of the following respiratory signals (resampled at 5Hz): R_{amp} , R_{pca} , R_{kpca} , R_{AB} or R_{TH} , with R_{AB} and R_{TH} the respiratory effort measured on the abdomen and on the thorax, respectively. All the information contained in X that is linearly related to Y , can be computed by projecting X onto a subspace \mathbb{V} defined by variations in Y . In order to define \mathbb{V} , three different approaches are implemented and described as follows.

- 1) The Wavelet decomposition of Y using *db4* as a mother wavelet and 5 levels is computed, as proposed in [23]. The corresponding approximation coefficients and its delayed version using m samples, are used to defined the basis \mathbf{V}_1 , describing the low frequencies of respiration, namely frequencies below 0.07Hz. On the other hand, the detail coefficients of level 3, 4 and 5, and their delayed versions using n samples, are used to construct the basis \mathbf{V}_2 , which describe respiratory variations between 0.07Hz and 0.6Hz.
- 2) The reference signal Y is first filtered using a band-pass, fifth order, Butterworth filter with cutoff frequencies at 0.04Hz and 0.15Hz. Next, the filtered signal and its delayed version are used to construct the basis \mathbf{V}_3 . Note that the range of frequency used here corresponds to the low frequency (LF) band commonly used in heart rate variability analysis [14]. In order to describe the high frequency (HF) band, Y is filtered using the same filter, but now with cutoff frequencies at 0.15Hz and 0.4Hz. This new filtered signal and its delayed version are used to construct \mathbf{V}_4 .
- 3) The subspace related to respiration is spanned by \mathbf{V}_5 , which is constructed using the original Y and its delayed version.

As mentioned in each approach, a delay is included in the different bases. This delay is computed using 10-fold cross-validation. The reason to use three approaches relies on the definition of the frequency bands. The first approach uses wavelet decomposition, which imposes frequency bands depending on the level of decomposition and the sampling

TABLE II
SET OF FEATURES. RESP CAN BE INTERPRETED AS ANY REAL OR ESTIMATED RESPIRATORY SIGNAL

Feature ID	Derived from			Frequency Range	Computation
	ECG	RR	Resp		
S_1		x		-	$std(RR)$
S_2			x	-	$std(R)$
r_k		x		-	$\frac{\sum_{i=1}^m (rr_i - \mu)(rr_{i+k} - \mu)}{\sum_{i=1}^m (rr_i - \mu)^2}$
PC	x			-	$\frac{\lambda_2}{\sum \lambda_i}$
F_1		x	x	0-0.07Hz	$F_i = \frac{x_i^T x_i}{X^T X},$ with
F_2		x	x	0.07-0.6Hz	
F_3		x	x	0.04-0.15Hz	$X_i = P_i X,$ $i = \{1, \dots, 5\},$ and
F_4		x	x	0.15-0.4Hz	
F_5		x	x	All	$X_6 = QX$
F_6		x	x	All	

frequency of the signal. Besides, in order to include a big part of the LF and HF bands this band configuration is used. In the second approach, the traditional LF and HF bands are explored. The last approach allows to describe “all” information of the heart rate that is related to respiration.

After constructing a basis V_i , with $i = \{1, \dots, 5\}$, a projection matrix

$$P_i = V_i(V_i^T V_i)^{-1} V_i^T \quad (1)$$

is used to find all the information of X contained in the subspace spanned by V_i , as

$$X_i = P_i X. \quad (2)$$

Note that X_i is a component of X related to variations of Y that are described by V_i . It is important to mention, that when computing X_5 , it is possible to describe all information contained in the heart rate that is *not* related to respiration. This will be denoted by X_6 , and can be computed as $X_6 = QX$, with $Q = I - P_5$.

Once X_i is obtained, its relative power is computed as

$$F_i = \frac{X_i^T X_i}{X^T X}, \quad (3)$$

and it corresponds to a feature that will be used as input to the classifier. In total, 6 features are computed for each ECG minute using orthogonal subspace projections, and they are summarized in Table II, which also includes all the features used in this study. These other features correspond to the standard deviation of the RR interval time series (S_1), standard deviation of the respiratory signal R (S_2), and the serial correlation coefficients of the RR (r_k). For the last one, $RR = \{rr_i\}_{i=1}^N$, with N the total amount of RR intervals per segment. μ corresponds to the mean of the RR intervals per segment, and $k = 1, \dots, 5$.

As mentioned before, the respiration modulates changes in the heart rate through the RSA, where the RR intervals are shortened during inspiration, and elongated during expiration. This mechanism allows for an improved gas exchange that leads to an optimal energy expenditure when respiration and

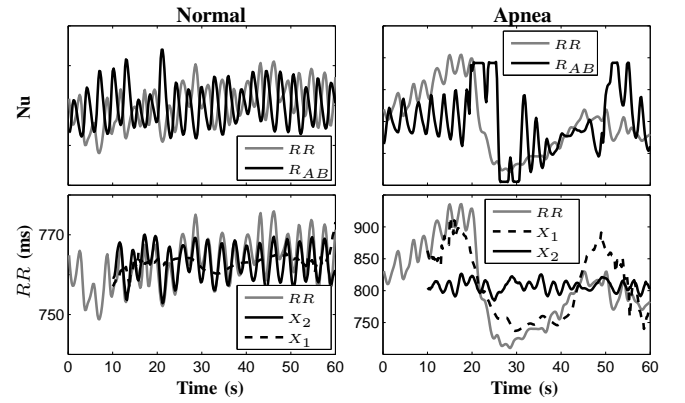


Fig. 4. (top) RR interval time series and respiratory signal measured in the abdomen for a normal (left) and an apneic (right) segment. Nu stands for normalized units. (bottom) Heart rate components derived using orthogonal subspace projections. X_1 corresponds to the component related to respiratory dynamics in the range between 0-0.07Hz, and X_2 describes variations in the range between 0.07-0.6Hz. Note that the components X_1 and X_2 get values from $t = 10s$ because the order of the model was 50, which corresponds to 10s for a sampling frequency of 5Hz.

heart rate are highly synchronized [21]. This synchronization is in some cases reduced during episodes of apnea [24]. However, a more important effect of apnea is related to an increased sympathetic modulation, which causes changes in heart rate that are not modulated by RSA [25]. The increased sympathetic activity has been already studied by means of the quantification of the sympathovagal balance, which can be computed as the ratio between the power of the heart rate in the low frequency band (0.04-0.15Hz) and the power in the high frequency band (0.15-0.4Hz). This ratio however, is defined in order to take into account *all* vagally and sympathetically mediated activities. Therefore, methodologies like the one described above, allow to determine whether or not respiration is the strongest modulator of the heart rate during a given minute epoch. This is to say, with this methodology, it is possible to extract the information of the heart rate that is modulated by mechanisms different from respiration. For instance, Fig 4 shows a normal minute of respiration and heart rate, and a minute containing obstructive sleep apnea. The two different heart rate components are also indicated in the figure. Note that the relative amplitude of the component related to respiration decreases during an episode of apnea.

G. Feature selection

In total, 28 features are extracted from each ECG minute, namely, the standard deviation of RR (S_1), the standard deviation of the three EDR signals (S_2), 5 serial correlation coefficients (r), the relative power of the second principal component (PC), and 18 features computed using orthogonal subspace projections (F). The reason to compute $18 \times F$ is that three EDR algorithms are used to estimate the respiration. In order to determine the discriminative power of each respiratory signal and of each feature, two different statistical tests are used, namely, the F-score [26] and the Kruskal-Wallis test.

The F-score for a given feature z_i , with $i = \{1, \dots, 28\}$, and the number of positive and negative instances denoted by

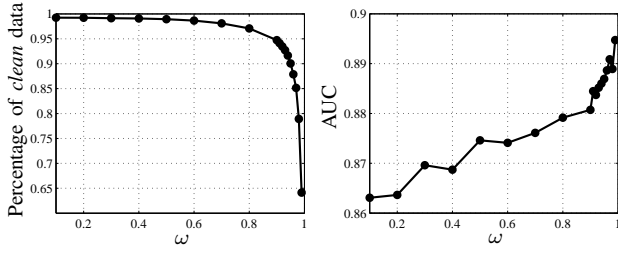


Fig. 5. (left) Percentage of the amount of data labeled as clean in the Physionet dataset with respect to the contamination level ω . (right) Area under the receiver operating characteristic curve (AUC) for different thresholds of ω . Note that when $\omega = 0.9$, about 95% of the data is considered as “clean”, and the slope of the performance changes.

n_+ and n_- , respectively, is computed as

$$F(i) \equiv \frac{(\bar{z}_i^{(+)} - \bar{z}_i)^2 + (\bar{z}_i^{(-)} - \bar{z}_i)^2}{Z_i^+ + Z_i^-}, \quad (4)$$

with

$$\begin{aligned} Z_i^+ &= \frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (z_{k,i}^{(+)} - \bar{z}_i^{(+)})^2 \\ Z_i^- &= \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (z_{k,i}^{(-)} - \bar{z}_i^{(-)})^2, \end{aligned} \quad (5)$$

where \bar{z}_i , $\bar{z}_i^{(+)}$, and $\bar{z}_i^{(-)}$ correspond to the average of the values of the i th feature for the whole set, the positive and the negative sets, respectively. The larger the F-score, the larger the discriminative power of the feature.

For each derived feature, one p -value (using the Kruskal-Wallis test), and one F-score are computed. In addition, these values are also computed for the transformed features obtained after applying the logarithm. This logarithmic transformation is used to obtain features with distributions closer to normal.

H. Classification

Once a set of features is selected for classification, segments with $\omega > 0.9$ are selected as candidates for the training set, which is then extracted using the fixed-size method [27]. The value of ω was selected using the values indicated in Fig. 5, where it is clear that when a threshold of 0.9 is applied on the values of ω , about 95% of the data is kept. In addition, the figure also shows that the slope of the curve AUC vs. ω changes for $\omega > 0.9$. Therefore, by selecting the training set from segments with $\omega > 0.9$, a maximum performance is achieved by only removing about 5% of the data.

This method finds a subset of data points that maximizes the Renyi entropy. In other words, the subset contains parameters derived from “clean” ECG segments, and it represents the underlying distribution of the dataset. This subset is used to perform training and validation, and the remaining data points are later used to test the apnea classifiers.

Different classifiers are used in this study, and they are listed as follows:

- 1) Linear Discriminant Analysis (LDA)
- 2) Support Vector Machines (SVM) using linear, polynomial and RBF kernels

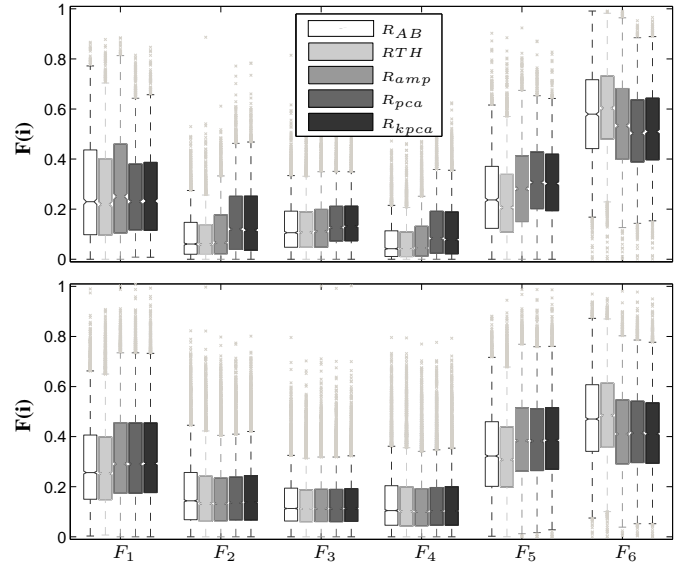


Fig. 6. Features extracted using orthogonal subspace projections in the Physionet (top) and Leuven (bottom) datasets. A comparison between the features obtained from the real respiratory signals and the three different EDR algorithms is indicated. No significant differences were found between the features. Note that the three EDR signals R_{amp} , R_{pca} and R_{kpca} , lead to very similar features.

- 3) Least-Squares SVM (LS-SVM) using linear, polynomial and RBF kernels [27].

Their performances are evaluated using sensitivity, specificity, area under the receiver operating characteristic (ROC) curve (AUC), and accuracy.

III. RESULTS

This section presents the comparison between the different EDR signals and the real respiratory efforts measured from the abdomen and thorax. A comparison between the different parameters extracted from either signal was performed for the two datasets separately, namely the Physionet, and the Leuven datasets. Furthermore, the features with the highest F -score and significant p -values were selected and will be presented below. In addition, different classifiers were evaluated, and a final algorithm to classify sleep apnea was identified. All experiments were carried out in MATLAB R2013a on an Intel(R) Core(TM) i7, 2.4GHz, 8GB RAM, running Windows 8.1.

A. Comparison between respiratory signals: real and EDR

In order to assess the usability of the EDR signals for the classification of sleep apnea, it is important to guarantee that the derived features do not differ significantly from those derived from the real respiratory signals. Hence, the features derived by means of orthogonal subspace projections on the heart rate and each respiratory signal (real or estimated) were compared. Fig. 6 shows the different relative powers F_i for each respiratory signal, and for the two datasets used in this study. The feature ID corresponds to the notation used in Table II. No significant differences were found between the respiratory signals. Moreover, it is clear from the figure that

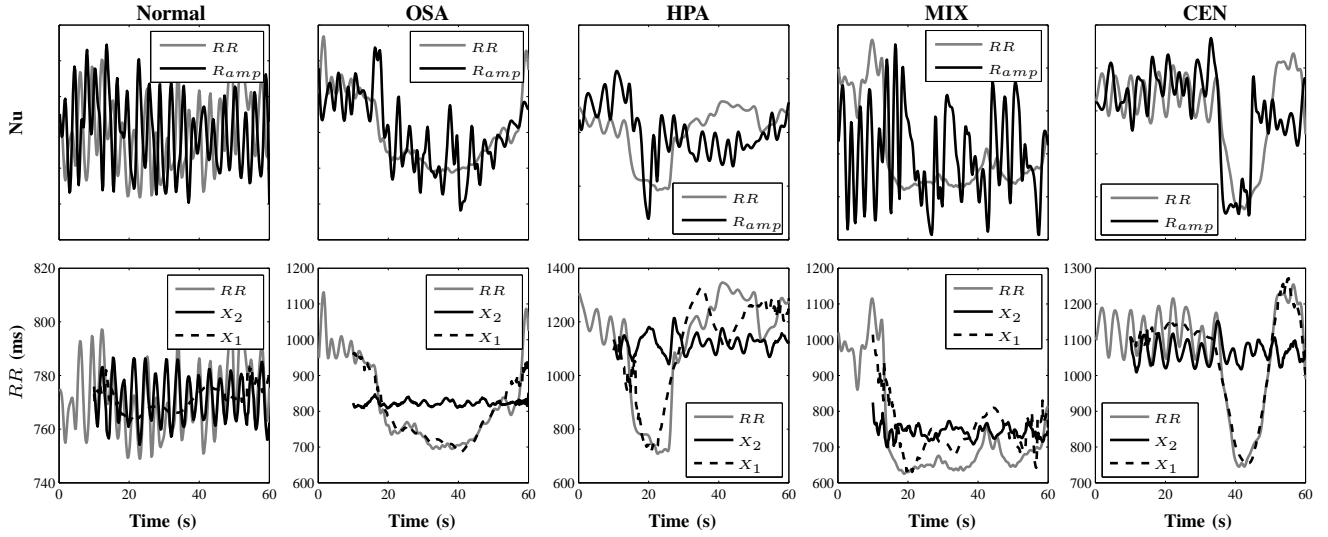


Fig. 7. Examples of the heart rate components derived using R_{amp} . Note that the dynamics of the EDR signal (R_{amp}) change in segments containing apnea events, which is an indication that the simplest EDR algorithm is sufficient to detect changes of the respiratory effort.

the three EDR signals led to the computation of very similar parameters.

It is important to mention that no discrimination between normal and apnea episodes was made at this point.

It can be observed from Fig. 6 that the respiratory information derived using the R-peak amplitude, R_{amp} , is similar to the one obtained with the real respiratory signals. Fig. 7 shows different examples of the components derived from the heart rate using R_{amp} , for a normal segment and different types of apneas. Note that in the segments containing apneas clear changes in the derived respiratory signal can be observed, and the amplitude of the components independent of respiration (dashed lines) become larger. This indicates that the simplest EDR algorithm already allows to observe variations in the respiratory effort, which are typical during apnea events.

B. Feature selection

After the observations described in the previous section it is clear that one can rely on features derived from estimated respiratory signals. This becomes of high importance, especially in datasets such as Physionet, where not all recordings include real measurements of respiration. With this in mind, only the features derived from the EDR were taken into account in the selection of the final feature set. In addition, no significant differences were found between different EDR. For this reason, the features computed from R_{amp} were favored due to the simplicity of the computation of the EDR using only the amplitude of the R-peaks. It was found that the features with F -scores larger than 0.3, and p -values lower than 0.05 for the distinction between apneas and normal segments were: F_1 , F_2 , PC , and the logarithms of S_1 , S_2 and r_5 . Remember that F_1 , F_2 and S_2 were computed using R_{amp} . Fig. 8 shows the different features for segments with and without apneas. Note that for the Leuven dataset, a discrimination between the different respiratory events is indicated. It is clear that the proposed features already provide a remarkable differentiation between respiratory events.

C. Apnea classification

In total, 34 324 ECG minutes were segmented on the Physionet dataset, and 5956 on the Leuven dataset. As mentioned in Section II-B, the similarity between the ACF is used to determine a level of contamination of the ECG segment. All segments with $\omega > 0.9$ were included in a subset of “clean” ECGs. From these clean ECGs (32 477 in Physionet, 5205 in Leuven)¹ training sets of 6000 segments from Physionet, and 3000 segments from the Leuven dataset, were selected by means of the fixed-size method. The remaining segments were used as test set and the comparison between classifiers was performed on this set.

Table III presents the performance for all tested classifiers, and for both datasets. It is clear from the table that the LS-SVM classifier using an RBF kernel outperforms the other classifiers in both datasets. Using the best classifier, the apnea-hypopnea index (AHI) is computed for each patient and each dataset. Results are shown in Fig. 9, where the real AHI is plotted against the estimated one. Note that the largest differences are obtained for larger AHI in both datasets. Some of these differences can be explained by the fact that apnea events can occur several times per minute, and since the data was segmented on a minute-by-minute basis, all those events are grouped into one single episode. If a threshold of 10min/h in apnea is applied to both datasets, it is possible to separate normal from apnea patients in the Physionet dataset, and patients with a higher than 10 AHI in the Leuven dataset, with an accuracy of 100%.

When looking at the classification results of the Leuven dataset, it is possible to manually identify the distribution of different types of respiratory events within the class labeled as *apnea*. This class contained 78.7% of 229 OSA, 37.7% of 496 OSH, 24.14% of 55 HPA, 93.48% of 48 MIX, 50% of 16 CEN, and 8.9% of 1454 normal segments, from the test set. As can be seen, if the OSH and HPA are excluded from the

¹In total, 1847 segments (5.3%) were removed from the Physionet dataset, and 751 (12.6%) from the Leuven dataset.

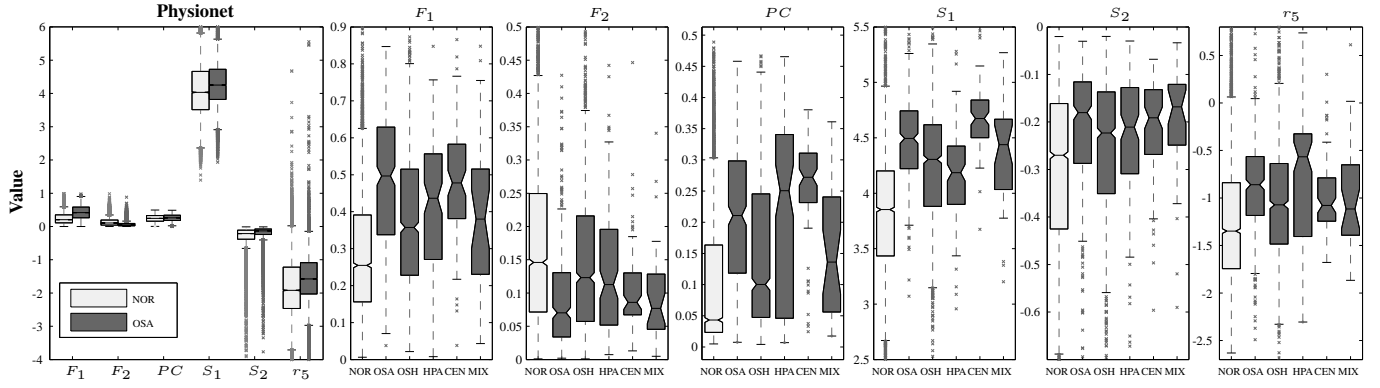


Fig. 8. Features selected as the most discriminative ones. The first panel in the left indicates the features for the physionet dataset, and the other 6 panels show the features for the Leuven dataset and for all respiratory events. The F -scores of the features are larger than 0.3 and their p -values are lower than 0.05 for the distinction between normal (NOR) segments and respiratory events. Note that the proposed features, F_1 , F_2 and PC , provide a clear differentiation between normal and obstructive apneas. Obstructive apneas: OSA; obstructive hypopneas: OSH; hypopnea: HPA; central apneas: CEN; and mixed apneas: MIX, are indicated.

TABLE III

PERFORMANCES OF THE TESTED CLASSIFIER ON THE PHYSIONET AND ON THE LEUVEN DATASETS. ALL RESPIRATORY FEATURES WERE COMPUTED USING R_{amp} . SENSITIVITY (SENS), SPECIFICITY (SPEC), ACCURACY (ACC), AND AREA UNDER THE ROC CURVE (AUC) ARE INDICATED.

Classifier		Physionet				Leuven			
		Sens.	Spec.	Acc	AUC	Sens.	Spec.	Acc	AUC
LDA		71.74%	71.20%	71.43%	78.46%	81.22%	79.71%	79.86%	87.69%
SVM	Linear	74.63%	68.51%	71.16%	78.5%	86.46%	74.6%	75.83%	87.74%
	Polynomial	78.36%	68.22%	72.6%	80.4%	88.21%	76.57%	77.78%	89.90%
	RBF	78.2%	70.55%	73.86%	81.85%	82.97%	81.78%	81.9%	88.41%
LS-SVM	Linear	71.48%	76.26%	71.78%	78.46%	80.35%	80.01%	80.05%	87.7%
	Polynomial	73.4%	73.43%	73.43%	81.18%	81.66%	81.7%	81.68%	89.17%
	RBF	84.71%	84.69%	84.74%	88.07%	78.81%	84.56%	83.95%	89.97%

analysis, the classifier achieves accuracies of more than 93%. These results will be discussed in the next section.

The results described above were obtained after selecting the training set using the “clean” ECG segments, which correspond to the segments with $\omega > 0.9$. If the training set is selected from the whole dataset, the AUCs using LS-SVM with an RBF kernel, are reduced to 80.2% and 84.1% for Physionet and Leuven test datasets, respectively. This shows that by including segments with a high confidence level in the training set, it is possible to reduce the bias in the classifier towards the presence of artifacts.

IV. DISCUSSION

This study had three main objectives, namely, the evaluation of the two novel features for the detection of sleep apnea, the comparison between three algorithms to derive respiratory information from the ECG, and the assessment of their added value in the classification of apnea. Bearing in mind the results presented in the previous section, it is clear that both features proposed here allow for a high discrimination between normal segments and respiratory events. Furthermore, the classification performance achieved with these two features is comparable with many algorithms proposed in the literature, where many more features are derived from the RR interval

time series and the EDR. For instance, for purely automatic algorithms based on single-lead ECG, accuracies of about 85 to 90% were reported in [6], [8] on the Physionet dataset. We believe that the difference in accuracy between this work and the literature is due to the size of the training set. In [6], [8], 35 recordings were used to train the classifiers, while in this work “only” 6000 ECG segments were selected as training set. The reason for this is that the algorithms based on SVM formulations require the computation of a kernel matrix, of which the size is limited by the hardware specifications. In this work, the size was limited to 6000×6000 . Nevertheless, even with a smaller training set, an accuracy of 84.74% was achieved for the Physionet dataset.

Concerning the feature derived from the morphology of the ECG, namely PC , it was observed that indeed, the relative power of the second component is higher due to variations in the QRS complexes caused by an increased sympathetic modulation.

Turning now to the feature derived using orthogonal subspace projections it is clear from the results that the information shared between heart rate and respiration, in the traditional LF and HF bands is significantly reduced during episodes of apnea. However, when looking at frequencies lower than 0.07Hz it was surprising to see that changes in respiration

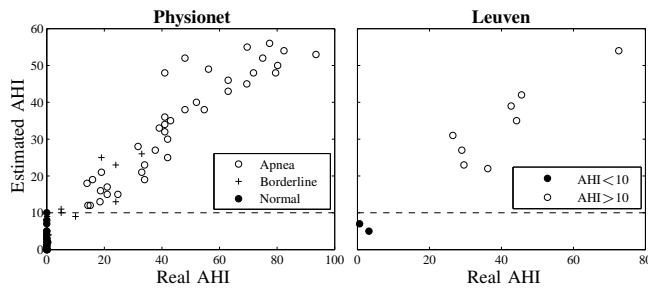


Fig. 9. Real and estimated apnea-hypopnea indices (AHI) for both datasets under investigation. Note that the maximum value for the estimated AHI is 60 due to the minute-by-minute analysis. A complete differentiation between the normal and apnea patients in the Physionet dataset can be made using a threshold of 10min/h (dashed line).

could also be observed in the heart rate. This could be the result of baseline variations in the respiration due to a reduced or complete absence of respiratory effort, which cause a rise in the heart rate at very low frequencies. It is important to remember that the analysis was performed on a minute-by-minute basis, which can be understood as a limitation in the frequencies that can be described. This is to say that only frequencies between 0.03Hz and 0.07Hz can be analyzed by this feature.

Another important point that requires attention is the fact that a comparison between three different algorithms to derive the EDR signal was performed. In addition, the features extracted from these signals were validated against those computed using the real respiration. It was remarkable to see how those features did not differ from each other. With this in mind it is clear that the usability of EDR signals on applications like sleep apnea detection is valid and significant. However, since no differences were observed between the three EDR algorithms, it is possible to rely on the simplest approach in order to reduce the complexity of the whole methodology. This last observation has clear advantages when developing portable devices, where less sensors are needed. Hence, if an accurate respiration can be estimated from the ECG, there is no need for respiratory belts.

It is not a surprise that the most difficult respiratory events to classify from the ECG signal are the hypopneas. As mentioned in the introduction, these events are associated with a reduced airflow and a reduced respiratory effort, and can be accompanied by a drop in oxygen saturation, or by a reduction of more than 50% of the flow during at least 10s, or by a reduced airflow with arousal or desaturation. Additionally, the difference between normal breathing and (central) hypopnea is not always clear [3]. This is because these respiratory events are not always accompanied by arousal, desaturation, snoring or thoraco-abdominal paradox. Therefore, they may appear very similar to normal segments. For this reason, in order to get a more reliable classification of this type of events the use of different physiological signals such as airflow, SpO_2 and CO_2 is required. Remember that the classifiers used in this work only discriminated between normal and respiratory events, and all apneas and hypopneas were considered to belong to one class labeled as *apnea*. The differentiation made before was done manually in order to determine the

distribution of respiratory events within the group classified as *apnea*.

V. CONCLUSIONS

A methodology for the automatic detection of sleep apnea from single-lead ECG was presented. Two novel features were proposed, one describes morphology changes of the QRS complexes caused by an increased sympathetic modulation during apnea, and one quantifies the information shared between respiration and HR at different frequency ranges. These two features together with two well-known parameters used in HRV, namely the standard deviation and the serial correlation coefficient allowed to discriminate apneas from normal activity. This discrimination was performed using an LS-SVM classifier, which proved to outperform other classifiers. The accuracies achieved in this study are comparable to the best results reported thus far for fully automated algorithms, namely in the range between 85% and 90%.

Apart from providing a label to each ECG minute, the methodology allows to estimate the contamination level of the recording by means of the autocorrelation function. This level was used in the training phase of the classifier, but it can also be used to determine the level of confidence on the results obtained for each segment. In other words, manual inspection of the highly contaminated segments can be recommended based on this feature.

Another contribution of this work was the comparison of three algorithms to derive respiratory information from the ECG. It was shown that the different parameters computed from the respiratory signals, real or estimated, were comparable for the detection of sleep apnea. Therefore, the use of only ECG sensors can guarantee up to 85% accuracies in the discrimination of both apneas and hypopneas together. In addition, these performances can be achieved using only the amplitude of the R-peaks. Hence, the added value of more elaborated EDR algorithms was negligible in this particular application. This is an advantage for home monitoring systems that require algorithms with low computational cost.

REFERENCES

- [1] P. E. Peppard et al., "Increased prevalence of sleep-disordered breathing in adults," *AM J Epidemiol*, vol. 177, no. 9, pp. 1006–1014, May 2013.
- [2] S. M. Caples et al., "Sleep-disordered breathing and cardiovascular risk," *Sleep*, vol. 30, no. 3, pp. 291–303, Mar 2007.
- [3] . Task Force AASM, "Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, no. 5, pp. 667–689, Aug 1999.
- [4] T. Penzel et al., "The apnea-ecg database," in *Proc. Computers in Cardiology*, 2000, pp. 255–258.
- [5] M. Bsoul et al., "Apnea medassist: real-time sleep apnea monitor using single-lead ecg," *IEEE Trans. Inf. Technol. B.*, vol. 15, no. 3, pp. 416–427, May 2011.
- [6] P. De Chazal et al., "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 686–696, Jun 2003.
- [7] K. Kesper et al., "Ecg signal analysis for the assessment of sleep-disordered breathing and sleep pattern," *Med. Biol. Eng. Comput.*, vol. 50, no. 2, pp. 135–144, Feb 2012.
- [8] T. Penzel et al., "Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings," *Med. Biol. Eng. Comput.*, vol. 40, no. 4, pp. 402–407, Jul 2002.

- [9] S. Boudaoud et al, "Corrected integral shape averaging applied to obstructive sleep apnea detection from the electrocardiogram," *EURASIP J. Adv. Sig. Pr.*, vol. 2007, no. 1, p. 032570, 2007.
- [10] P. Langley et al, "Principal component analysis as a tool for analyzing beat-to-beat changes in ecg features: application to ecg-derived respiration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 821–829, Apr 2010.
- [11] D. Widjaja et al, "Application of kernel principal component analysis for single-lead-ecg-derived respiration," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 1169–1176, Apr 2012.
- [12] A. L. Goldberger et al, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun 2000.
- [13] P. de Chazal et al, "Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram," *Physiol. Meas.*, vol. 25, no. 4, p. 967, Aug 2004.
- [14] Task Force ESC and NASPE, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar 1996.
- [15] C. Varon et al, "Robust artefact detection in long-term ecg recordings based on autocorrelation function similarity and percentile analysis," in *Proc. IEEE Eng. Med. Bio.*, 2012, pp. 3151–3154.
- [16] R. J. Thomas et al, "Differentiating obstructive from central and complex sleep apnea using an automated electrocardiogram-based method," *Sleep*, vol. 30, no. 12, p. 1756, Dec 2007.
- [17] A. H. Khandoker et al, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings," *IEEE Trans. Inf. Technol. B.*, vol. 13, no. 1, pp. 37–48, Jan 2009.
- [18] G. B. Moody et al, "Derivation of respiratory signals from multi-lead ecgs," *Proc. Computers in cardiology*, vol. 12, no. 1985, pp. 113–116, 1985.
- [19] G. D. Clifford et al, *Advanced methods and tools for ECG data analysis*. Artech House, Inc., 2006.
- [20] J.-Y. Kwok and I. W. Tsang, "The pre-image problem in kernel methods," *IEEE Trans. Neural Networ.*, vol. 15, no. 6, pp. 1517–1525, Nov 2004.
- [21] G. G. Berntson et al, "Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications," *Psychophysiology*, vol. 30, no. 2, pp. 183–196, Mar 1993.
- [22] A. Caicedo et al, "Cardiopulmonary coupling in sleep apnea assessed by means of orthogonal subspace projection," in *Proc. the 5th International Conference of Advances in Medical Signal and Information Processing (MEDSIP)*, 2012.
- [23] —, "Decoupling the influence of systemic variables in the peripheral and cerebral haemodynamics during ecmo procedure by means of oblique and orthogonal subspace projections," in *Proc. IEEE Eng. Med. Bio.*, 2012, pp. 6153–6156.
- [24] T. Penzel et al, "Cardiovascular and respiratory dynamics during normal and pathological sleep," *Chaos*, vol. 17, no. 1, p. 015116, Mar 2007.
- [25] E. Vanninen et al, "Cardiac sympathovagal balance during sleep apnea episodes," *Clin. Physiol.*, vol. 16, no. 3, pp. 209–216, May 1996.
- [26] Y.-W. Chen and C.-J. Lin, "Combining svms with various feature selection strategies," in *Feature extraction*. Springer, 2006, pp. 315–324.
- [27] J. A. K. Suykens et al, *Least squares support vector machines*. World Scientific, 2002, vol. 4.



Carolina Varon (M11) received the degree in electronic engineering from the Universidad de Ibagué, Colombia, in 2005, the M.Sc. degree in astronomy and astrophysics in 2009 and the M.Sc. degree in artificial intelligence in 2010, both from the KU Leuven, Belgium, where she is currently working towards the Ph.D. degree in the Department of Electrical Engineering. From 2002 to 2003, she was a Site Engineer in Agroindustrial del Tolima, Ibagué, Colombia. From 2005 to 2007, she was the Technical Support for Latin America for Security

Solutions, Bogotá, Colombia. Her research interests include biomedical signal processing, machine learning, and kernel methods.



in KU Leuven, Belgium. His research interests include biomedical signal processing, kernel methods, system analysis and control theory.



of sleep-related breathing disorders, non-invasive ventilation and respiratory muscle function.



plication techniques.

Alexander Caicedo received the degree in electronic engineering in 2005 and the M.Sc. degree in control engineering in 2007, both from the Universidad de Ibagué, Colombia. From 2007 to 2009, he worked as an assistance professor at Universidad de Ibagué, Colombia. He obtained his Ph.D. degree in Engineering from the KU Leuven, Belgium in 2013. During his PhD he developed several algorithms for the assessment of cerebral autoregulation in premature infants. He is currently a postdoctoral research at the department of electrical engineering (ESAT)

Dries Testelmans graduated in Medicine and Pneumology at the University Hospitals Leuven, Belgium. He performed his PhD at the Department of Respiratory Diseases of the KU Leuven. He further specialized in Sleep Medicine at the sleep lab of the Centre Hospitalier Universitaire in Grenoble, France. Currently, he is adjunct head of clinic at the Department of Respiratory diseases and at the Leuven University Centre for Sleep/Wake Disorders of the University Hospitals Leuven. Besides clinical practice he is also performing research in the field

Bertien Buyse graduated in Medicine and Pneumology at the University Hospitals Leuven, Belgium. She performed her PhD at the Department of Respiratory Diseases of the KU Leuven. Currently, she is head of clinic at the Department of Respiratory diseases and director at the Leuven University Center for Sleep/Wake Disorders of the University Hospitals Leuven. Besides clinical practice she is also performing research in the field of sleep-related breathing disorders, regulation of breathing, non-invasive ventilation and diaphragmatic pacing and



Special attention is given to the numerical aspects and to the design of reliable algorithms and their practical evaluation in medical diagnostics.

Sabine Van Huffel (M96A96SM99F09) received the M.Sc. degree in computer science engineering, the M.Sc. degree in biomedical engineering, and the Ph.D. degree in electrical engineering from the Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium, in June 1981, July 1985, and June 1987, respectively. She is currently a Full Professor in the Department of Electrical Engineering, KU Leuven. Her research interests include numerical (multi)linear algebra and software, system identification, parameter estimation, and biomedical data processing.