Capacitively-coupled ECG and Respiration for the Unobtrusive Detection of Sleep Apnea

Margot Deviaene[1,2]**, Ivan D. Castro**[3]**, Pascal Borzée**[4]**, Aakash Patel**[3]**, Tom Torfs**[3]**, Bertien Buyse**[4]**, Dries Testelmans**[4]**, Sabine Van Huffel**[1,2]**, and Carolina Varon**[1,2]

[1] Department of Electrical Engineering-ESAT, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven B-3001, Belgium
[2] Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium.
[3] Imec, Leuven B-3001, Belgium.
[4] Department of Pneumology, Leuven University Centre for Sleep and Wake Disorders, UZ Leuven, Leuven 3000, Belgium.

Abstract.
Objective: The performance of a novel unobtrusive system based on capacitively-coupled electrocardiography (ccECG) combined with different respiratory measurements is evaluated for the detection of sleep apnea. *Approach:* A sleep apnea detection algorithm is proposed, which can be applied to electrocardiography (ECG) and ccECG, combined with different unobtrusive respiratory measurements, including ECG derived respiration (EDR), respiratory effort measured using the thoracic belt (TB) and capacitively-coupled bioimpedance (ccBioz). Several ECG, respiratory and cardiorespiratory features were defined, of which the most relevant ones were identified using a random forest based backwards wrapper. Using this relevant feature set, a least-squares support vector machine classifier was trained to decide if a one minute segment is apneic or not, based on the annotated polysomnography (PSG) data of 218 patients suspected of having sleep apnea. The obtained classifier was then tested on the PSG and capacitively-coupled data of 28 different patients. *Main results:* On the PSG data, an AUC of 76.3 % was obtained when the ECG was combined with the EDR. Replacing the EDR with the TB led to an AUC of 80.0 %. Using the ccECG and ccBioz or the ccECG and TB resulted in similar performances as on the PSG data, while using the ccECG and ccECG-based EDR resulted in a drop in AUC to 67.4 %. *Significance:* This is the first study which tests an apnea detection algorithm on capacitively-coupled ECG and bioimpedance signals and shows promising results on the capacitively-coupled data set. However, it was shown that the EDR could not be accurately estimated from the ccECG signals. Further research into the effect that respiration has on the ccECG is needed to propose alternative EDR estimates.

Keywords Apnea detection, capacitively-coupled ECG, Bioimpedance

1. Introduction

The most common sleep related breathing disorder is obstructive sleep apnea (OSA), which presents itself as partial or complete cessations of breathing during sleep. It is estimated that worldwide, almost 1 billion people suffer from OSA, most of them, however, remain undiagnosed and consequently untreated (Benjafield et al., 2019). This is partly due to the expensive and cumbersome way of diagnosing by means of a polysomnography (PSG), which is a diagnostic test usually performed using an overnight stay in the hospital during which several physiological parameters are measured. Based on this PSG, sleep stages and respiratory events are afterwards manually scored by a sleep expert according to a set of scoring rules (Berry et al, 2012).

Several studies have been performed to develop home-monitoring systems for the detection of sleep apnea in order to enable screening on a larger scale, and to follow-up patients on the longer term. This long-term follow-up can be useful to obtain a quantitative assessment of the effect of different OSA treatments and lifestyle changes (Adult Obstructive Sleep Apnea Task Force of the AASM, 2009). Respiratory sensors used during the PSG, such as the nasal pressure sensor and the thoracic and abdominal respiratory effort belts have been used to automatically detect OSA (Mendonca et al., 2018; Uddin et al., 2018). Apneas have been shown to cause secondary responses, such as bradycardia, which is followed by tachycardia when the breathing is restored (Guilleminault et al., 1984). Moreover, the heart rate variability is modulated by the respiration by means of the respiratory sinus arrhythmia, and this modulation is attenuated during apneic events (Varon et al., 2015a). This, combined with the development of several methods to extract an ECG derived respiration (EDR) (Varon et al., 2020), has led to the development of multiple ECG-based OSA detection algorithms (Faust et al., 2016; Mendonca et al., 2018; Varon et al., 2015a).

Although using only an ECG patch for the detection of sleep apnea is already much more comfortable for the patient than the full PSG, electrode-skin-contact is still required, which makes it not ideal for long-term monitoring (Spinelli and Haberman, 2010). As a consequence, non-contact ECG sensors have been developed. In this study, a capacitively-coupled ECG (ccECG) sensor, integrated in the mattress, was used (Castro et al., 2019b). In this sensor, the galvanic contact between the skin and the electrode is replaced by a capacitive coupling between the skin and the conductive plate of the sensor. All non-conductive materials in between, such as the bed sheets and clothing, serve as dielectric. The sensor also includes a capacitively-coupled bioimpedance (ccBioz) measurement, which can be used to extract the respiratory signal (Castro et al., 2019b).

The goal of this study was to develop an automated apnea detection algorithm which can be applied on the (cc)ECG signals combined with different respiratory measurements such as the thoracic belt (TB) and unobtrusive tools: an EDR estimate, and the ccBioz. The algorithm was developed based on the ECG and TB, extracted from the PSG. In the testing phase, these gold-standard signals were replaced by their capacitively-coupled alternatives. As such, the usability and limitations of ccECG and

ccBioz measurements for the automatic detection of sleep apnea were explored. To our knowledge, this is the first study investigating automatic OSA detection based on ccECG and ccBioz recordings. This study builds upon the work described in (Varon et al., 2015a) and (Castro et al., 2018). Here, additional cardiorespiratory measures were investigated and features extracted from the gold-standard PSG signals and the capacitively-coupled measurements were compared. A set of relevant features were selected with and without taking into account if a feature behaves similar when extracted from the PSG and capacitively-coupled signals. Furthermore, the use of a state-of-the-art EDR signal was evaluated for the first time on ccECG.

In the next sections, the data sets will be discussed, followed with the data preprocessing including heart rate and EDR extraction. Subsequently, the feature extraction, selection and classification will be explained. In Section 4 the results will be presented, which will be further discussed in Section 5.

2. Materials

This study includes three data sets of gold standard in-hospital PSGs, recorded from subjects suspected of suffering from OSA at the sleep laboratory of the University Hospitals Leuven (UZ Leuven). The inclusion of these data sets was approved by the ethical committee of UZ Leuven (S53746, S60319) and all patients signed an informed consent. An overview of the patient demographics of each data set can be found in Table 1.

The first data set, denoted *OSA_ref*, contains 218 PSGs and was used to train the classifier. The second data set, called *OSA_cc*, is used to test the classifier and consists of 28 PSG recordings, for which ccECG and ccBioz signals were simultaneously measured using an array of sensors embedded in the mattress. This sensor configuration was provided by Imec, and it is described in (Castro et al., 2018) and (Castro et al., 2019b). An overview of the recording set-up is shown in Figure 1. Due to technical problems (i.e. unidentified electrode disconnection in the prototype), the simultaneous ccBioz signal could only be collected for 11 subjects. The third data set, used to test

Table 1: Patient demographics of the included data sets

Data set	# Sub.	Age Years	BMI Kg/m^2	AHI Events/h	AHI $\geq$15	Male
OSA_ref	218	49±12 (41, 58)	30±6 (25, 34)	27±24 (8, 39)	58%	65%
OSA_cc	28	48±14 (38, 55)	30±5 (25, 35)	28±26 (6, 58)	57%	50%
OSA_Varon	10	48±11 (43, 54)	30±5 (27, 31)	33±20 (27, 44)	80%	90%

Age, BMI and AHI are presented as mean ± standard deviation, with the 25 % and 75% quantile values underneath.
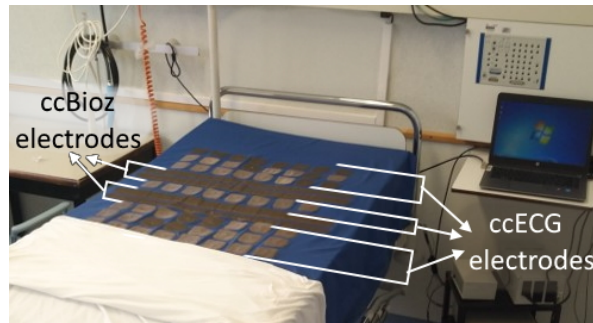
Figure 1: Recording setup of the mattress with capacitively-coupled sensors embedded, measuring ccECG and ccBioz.

the classifier, corresponds to the data set used in (Varon et al., 2015a), which contains 10 PSGs. This data set is referred to as *OSA_Varon*.

An overview of the recording characteristics for each of the data sets is given in Table 2. All data sets were manually annotated by sleep specialists according to the AASM 2012 scoring rules (Berry et al, 2012) or the Chicago criteria (Quan et al., 1999). When the AASM 2012 scoring rules are followed, apneas are scored when the airflow amplitude decreases more than 90 %. Hypopneas, on the other hand, only need a decrease of at least 30 %, accompanied by either an arousal or an oxygen desaturation of at least 3 %. Whereas, according to the Chicago criteria, events are scored when either an airflow drop larger than 50 % occurs, or when a clear airflow amplitude reduction is observed accompanied by an arousal or an oxygen desaturation of at least 3 %. Both scoring rules require events to last at least 10 seconds.

These annotations were provided together with the PSG ECG and thoracic respiratory effort belt (TB), recorded using a respiratory inductance plethysmography belt. The PSG signals were sampled at 500 Hz, or at 200 Hz for the older *OSA_Varon* data set, whereas the capacitive sensor recorded at 512 Hz. The PSG signals were, therefore, upsampled to 512 Hz in order to match the capacitive sensor's sampling frequency. The signals from both measurement systems were synchronized using the tachograms extracted from the PSG ECG and ccECG.

Table 2: Overview of the recording characteristics for each data set. The used apnea scoring rules, the sampling frequency (Fs) for the PSG and capacitive sensor recordings, and number of available recordings per channel are denoted.

Data set	Scoring rules	PSG			Capacitive sensor		
		Fs (Hz)	# ECG	# TB	Fs (Hz)	# ccECG	# ccBioz
OSA_ref	AASM 2012	500	218	218	-	0	0
OSA_cc	AASM 2012	500	28	28	512	28	11
OSA_Varon	Chicago	200	10	0	-	0	0

3. Methods

Before analyzing the signals, the data was preprocessed in order to remove noise and artefacts. Then, the EDR and RR-interval time series were extracted. From these signals, an apnea detection classifier was developed based on the PSG data; features were computed of which the most relevant ones were selected and fed to a classifier. This model was then tested on the capacitively-coupled signals and additional experiments were performed in order to further analyze the obtained results.

3.1. Data preprocessing

An overview of the preprocessing steps needed to prepare the different signals for feature extraction is given in Figure 2. These steps will be discussed in more detail below. After signal prefiltering, all signals were split into 1-minute windows and the further analysis was performed on a window base.

3.1.1. ECG signals

ECG prefiltering: A band pass butterworth filter was applied using cutoff frequencies of 0.67 and 40 Hz (Castro et al., 2019b, 2018). The used low-pass cut-off frequency of 40 Hz is lower than what is recommended by the task force of heart rate variability (Malik, 1996). But, when this upper limit is increased for unobtrusive ECG measurement systems, a significant increase in noise levels is observed. Therefore, the IEC 60601-2-47 standard for ambulatory ECG systems was followed to define the filtering band.

      Check for signal inversion: A check for signal inversion needed to be implemented since for the PSG signals the placement of the ECG electrodes was sometimes adapted, and the ccECG array sensor will select every minute the best two electrodes to record the ccECG (Castro et al., 2019b). Therefore, signal inversion can
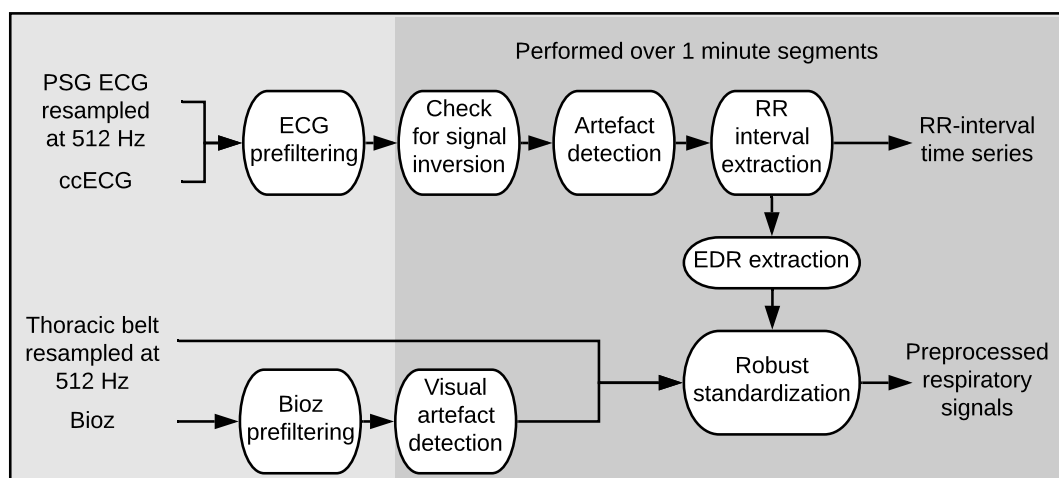

Figure 2: Overview of the applied signal preprocessing methods.

occur between segments for the ccECG. When the absolute value of the 1st percentile value was larger than the 99th percentile value, lead inversion was assumed and the signal was inverted.

Artefact detection: For each 1-minute window, a signal quality score was computed. For the PSG ECG, this was done based on the auto correlation function of the ECG (Moeyersons et al., 2019). Segments with a quality score lower than 90 were discarded. Whereas for the ccECG, the algorithm proposed by Castro et al. was applied (Castro et al., 2019a). In this case, only the high quality data segments were included using the high-threshold linear SVM model from (Castro et al., 2019a).

RR-interval extraction: The R-peaks were extracted from the ECG segments using the algorithm discussed in (Varon et al., 2015a), which makes use of the upper and lower envelopes to flatten the ECG signal. Based on the location of the detected R-peaks, the RR-interval time series, or tachogram was derived.

EDR extraction: From each 1-minute window, where at least 30 R-peaks were detected, the EDR was computed. Two EDR algorithms were compared: the R-wave amplitude (Ramp) and the downward slope of the the R-wave (RS slope) (Lázaro et al., 2014). These were extracted as described in (Varon et al., 2020).

3.1.2. Respiratory signals

ccBioz prefiltering: The bioimpedance signals were lowpass filtered at 1 Hz. The PSG TB signals were clean and did not require any additional prefiltering.

Visual artefact detection: No artefact removal was necessary for the TB respiratory signal, as we assumed movement artefacts overlapped with the ones detected in the PSG ECG. For the ccBioz signals, on the other hand, a visual quality check was performed per 1-minute window and noisy segments were removed.

Robust standardization: In order to ensure that features extracted from the different respiratory signals behave similarly, a robust standardization was applied to all respiratory signals, including the EDRs. The data was rescaled such that the 10th and the 90th percentile values of the entire signal corresponded to values of, respectively, -1 and 1.

As a last preprocessing step, the continuous apnea annotations of the PSG were converted to 1-minute labels. Whenever a 1-minute segment overlapped with an annotated event (including both apneas and hypopneas), the segment was labeled as apneic. These 1-minute labels were used as reference to train and test the classifier.

An overview of the available number of clean hours of data per data set and signal modality can be found in Table 3.

3.2. Training apnea detection classifier

The apnea detection classifier was trained on the preprocessed PSG signals of the *OSA_ref* data set. The training workflow is depicted in Figure 3 (top). Afterwards, the classifier was tested on different combinations of ECG and respiratory measurements

Table 3: Number of hours of clean data for each data set and each signal modality.

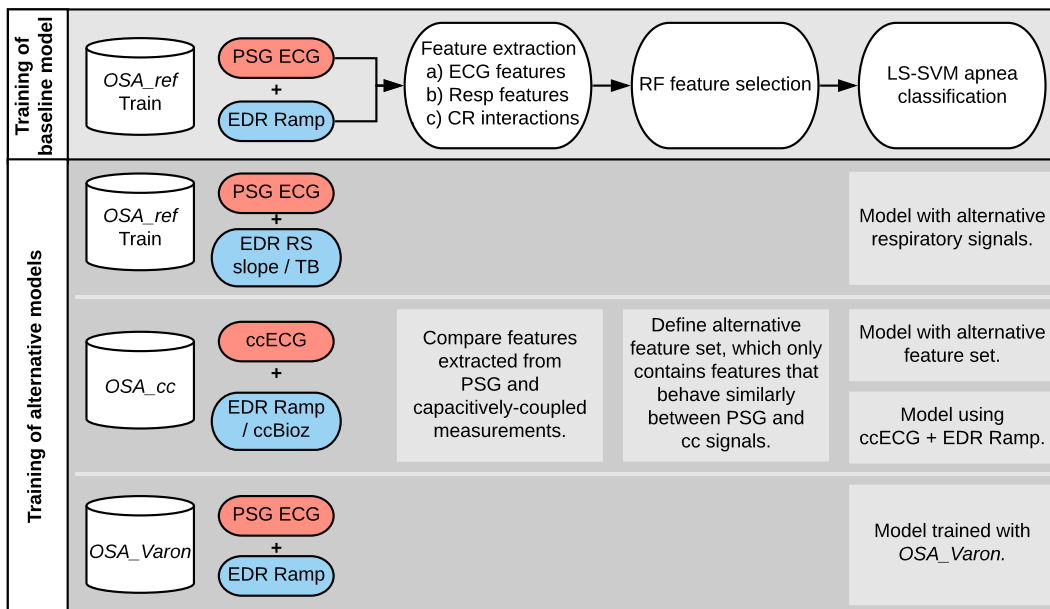|		Hours of clean data Normal sleep/Apneic			
		PSG		Capacitive sensor	
Data set	Total hours of recording Normal sleep/Apneic	ECG	TB	ccECG	ccBioz
OSA_cc	255 172/83	193 128/65	193 128/65	74 52/22	19 14/5
OSA_ref	1990 1381/609	1547 1086/461	1547 1086/461	NA	NA
OSA_Varon	99 64/35	82 55/27	NA	NA	NA

NA: Not available.

Figure 3: Development of the apnea detection classifier. The top row depicts the different steps used to train the classifier. Below, an overview is given of the additional experiments conducted to further analyze the obtained results.

of the PSG and capacitively-coupled data sets, as shown in Figure 4, and additional experiments were conducted in order to explain the obtained results (see Figure 3). The different steps in the development of the classifier are explained below. The additional experiments will be discussed in the next section.

3.2.1. Feature extraction

ECG features: A total of 27 features were extracted from the PSG ECG and ccECG segments. The mean, standard deviation (std) and kurtosis of the RR-intervals were extracted, as well as the SDSD (standard deviation of RR-interval differences) and the RMSSD (square root of the mean squared differences of successive RR-intervals), the
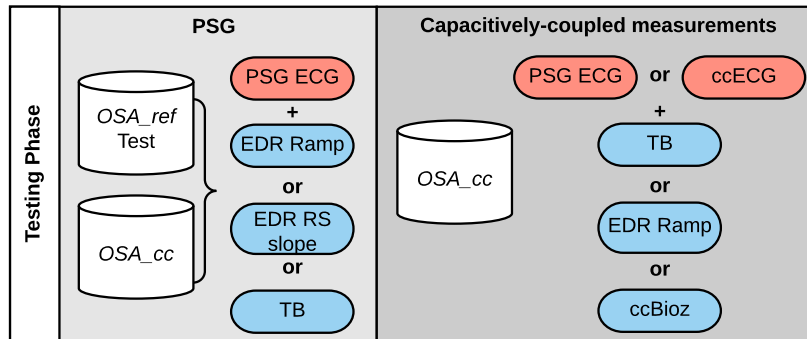
Figure 4: Signal modality combinations per data set on which the classifier was tested.

first five serial correlation coefficients of the RR-intervals and the Allan factor (on time scales of 5, 10, 15, 20, 25 and 30 s) (De Chazal et al., 2003).

Additionally, the QRS$_{\text{morph}}$ feature quantifies changes in the QRS morphology. This feature was computed by applying principal component analysis on a matrix composed of the aligned QRS complexes. QRS$_{\text{morph}}$ was taken as the percentage of variance explained by the second component (Varon et al., 2015a).

Also, the following heart rate variability frequency parameters were included: the power in the high frequency (0.15-0.4 Hz, HF) and low frequency (0.04-0.15 Hz, LF) bands, their normalized versions, and the LF/HF.

Finally, the phase rectified signal averaged (PRSA) downward curve of the RR-intervals was computed with a window length of 10 seconds (Bauer et al., 2006; Varon et al., 2015b). From this PRSA curve, the amplitude difference between the highest and lowest peaks around the anchor point, overall slope, slope before and after the anchor point, and the accelerating capacity were extracted as features.

Respiratory features: From the different respiratory signals, 13 features were extracted: the mean, standard deviation and kurtosis; the five PRSA features, described above, extracted from the downward PRSA curve of the respiratory signal with a window length of 10 seconds; and the power of the signal in the HF and LF bands, their normalized versions and LF/HF.

Cardiorespiratory interactions: A large number of features to estimate the cardiorespiratory coupling were investigated. These were computed on each combination of an ECG and a respiratory signal. An overview of the considered signal modality combinations can be found in Figure 4. Cardiorespiratory interactions based on the respiratory bandwidth, information dynamics, bivariate PRSA, time frequency representation and subspace projections were computed (Morales et al., 2019).

Additionally, the respiratory and RR-interval signals were decomposed in five db4 wavelet levels, from which the respiratory component of the heart rate was extracted using subspace projections as introduced in (Varon et al., 2015a). The low frequency respiratory variations (WAV$_{\text{LF}}$, 0 - 0.07 Hz) and high frequency respiratory variations (WAV$_{\text{HF}}$, 0.07-0.6 Hz) were extracted.

3.2.2. Feature selection A total of 117 features were extracted from the different signals, the logarithm was also computed for all of them and considered as feature. In order to find the most relevant features for the detection of sleep apnea, feature selection was performed on the *OSA_ref* set.

A random forest based wrapper was implemented (Deviaene et al., 2019). For each iteration, the feature whose removal resulted in the lowest drop in Cohen kappa value on the patient-independent out-of-bag samples, was deleted from the feature set. In order to speed up the computations, highly correlated features were removed on beforehand. From each feature pair with a correlation coefficient larger than 0.9, only the feature with the largest F-score was retained (Deviaene et al., 2018).

3.2.3. Classification Once the relevant features were discovered, a classifier was trained to predict whether or not a 1-minute segment contains an apnea. A fixed-size least-squares support vector machine (LS-SVM) classifier with RBF kernel was trained based on the *OSA_ref* data set. The 6000 most representative samples were selected for training based on the entropy criterion (De Brabanter et al., 2002). The hyperparameters of the LS-SVM model were tuned using the *tunelssvm* function. 10-fold cross-validation was performed on the training set using the misclassification rate as cost function. The performance of the model was computed on the remaining samples of *OSA_ref* and the two other data sets, using the area under the receiving operating characteristic (ROC) curve (AUC), the accuracy (Acc), sensitivity (Se) and specificity (Sp). The optimal cutoff point on the ROC was selected based on the optimization of the multiplication of Se and Sp. An overview of the different signal modality combinations on which the developed algorithm was tested for each data set can be found in Figure 4. For the PSG data, the performance results were compared when different types of respiratory measurements were used. On *OSA_cc*, the influence on the performance was studied when the PSG signals were replaced by the capacitively-coupled measurements. For *OSA_Varon*, only the combination of PSG ECG and EDR Ramp was considered, for which the performance was compared against the results from (Varon et al., 2015a).

3.3. Additional experiments

In order to further investigate the results, additional experiments were conducted. An overview of these experiments can be found in Figure 3. Alternative models were obtained by retraining the LS-SVM classifier from scratch with different training sets or signal modalities, no transfer learning was applied.

3.3.1. Alternative respiratory signals To improve the comparison between the use of different PSG respiratory modalities, an alternative classifier model was trained with features computed from the PSG ECG combined with EDR RS slope and TB for *OSA_ref* train.

3.3.2.   Comparison of features extracted from PSG and capacitively-coupled measurements The feature selection is solely based on the PSG based features from the training data set. But, as concluded in (Castro et al., 2018), not all ECG based features can be extracted accurately from the ccECG signals. The obtained features might therefore not be the most relevant ones for the ccECG. In order to investigate the relationship between features extracted from the ccECG and PSG ECG, a linear regression was performed in which the ccECG features were used as regressors and the PSG features as target values. The R-squared values of these regressions were used to evaluate how well the PSG features could be extracted from the capacitively-coupled signals.

An additional feature selection was then performed in which only features which had an R-squared value above a certain threshold were considered. Based on this new feature set, an alternative classification model was trained using the PSG data of *OSA_ref*.

3.3.3.   Classification model based on capacitively-coupled data In order to further optimize the classification model towards the capacitively-coupled measurements, the model was trained based on the ccECG and EDR Ramp data of *OSA_cc*. Unfortunately, not enough ccBioz data was available to be included in this experiment.

10-fold cross validation was performed on the patients included in this data set. The patients were ranked according to their AHI. Every group of 10 patients in this ranking was then spread over the 10 folds, as such, each fold included a wide range of OSA severity levels.

3.3.4. Training model using OSA_Varon The apnea scoring rules have changed since the *OSA_Varon* data set was recorded, and moreover, the patient demographics differ between *OSA_Varon* and *OSA_Ref*. Therefore, a model trained on the latter data set might not generalize well on the former. In order to ensure a fair comparison to the results obtained in (Varon et al., 2015a), a model was retrained on a subset of *OSA_Varon*.

4. Results

4.1. Extracted features

Seven features were identified to be the most relevant ones using the *OSA_ref* data set. These include:

- two ECG features: the log standard deviation of the RR-intervals and QRS_{morph}.
- two respiratory features: Log standard deviation and log kurtosis of the respiration.
- three cardiorespiratory interaction features: WAV_{LF}, log WAV_{HF} and the log 3-dB bandwidth of the respiratory power spectral density estimate (log PSD BW resp).

In Figures 5 and 6, four of the selected features are plotted for different combinations of ECG and respiratory measurements for the *OSA_cc* data set. In Figure 5, the
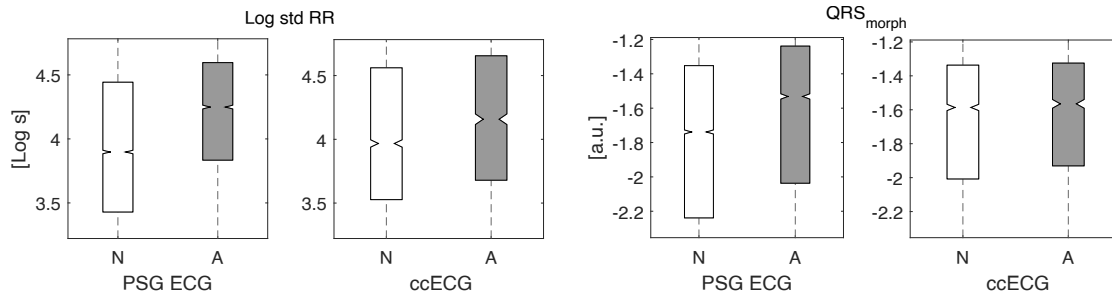
Figure 5: ECG features extracted from OSA_cc for normal (N) and apneic (A) segments using the PSG ECG and ccECG.
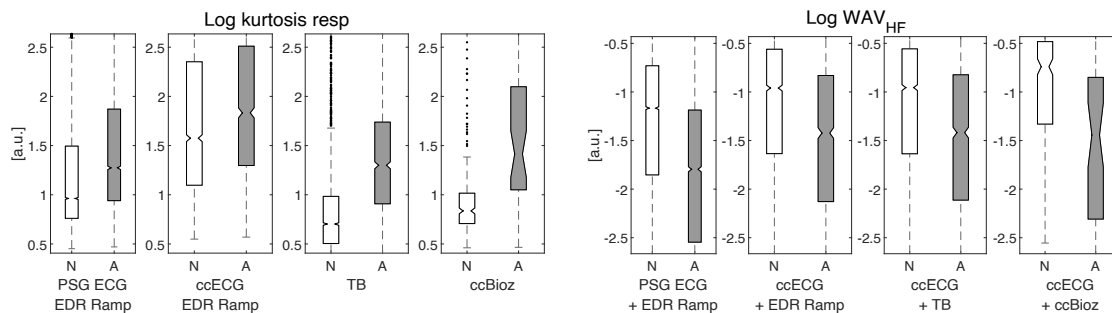

Figure 6: Log kurtosis of the respiration and Log WAV$_{\text{HF}}$ feature extracted from OSA_cc for normal (N) and apneic (A) segments using different signal modality combinations.

QRS$_{\text{morph}}$ and log std RR features are compared when extracted from the PSG ECG and ccECG. It is clear that the separation between normal (N) and apneic (A) segments is better for the PSG ECG signals than for the ccECG, especially when using a morphology-related feature, as suggested in (Castro et al., 2018). Figure 6 depicts the log kurtosis of the different respiratory measures. The TB and ccBioz respiratory signals show a better separation than when EDR signals are used. The selected cardiorespiratory interaction features, on the other hand, show a similar separation for all signal modality combinations, as can be seen for log WAV$_{\text{HF}}$ in Figure 6.

Using the Kruskal-Wallis test, the features in Figures 5 and 6 were shown to be statistically different between normal sleep and apneic segments with a significance level of 0.05. The QRS$_{\text{morph}}$ feature, however, had the highest p values with a value of 0.03 when using the ccECG.

4.2. Apnea detection

Using the seven extracted features, three classifiers were trained using the OSA_ref PSG ECG combined with the different respiratory measurements (EDR Ramp, EDR RS slope and TB). The performance results on the test set of OSA_ref and the PSG data of OSA_cc can be found in Table 4. The performances on the OSA_cc data set are slightly better than on the OSA_ref test set. Using the TB an AUC of 80 % is obtained

Table 4: Apneic segment classification performance on the PSG data, trained on the corresponding signal combination.

| | *OSA_ref test* | | | *OSA_cc* | | |
	EDR Ramp	EDR RS slope	TB	EDR Ramp	EDR RS slope	TB
AUC [%]	73.5	73.3	79.5	76.3	75.1	**80.0**
Acc [%]	68.8	68.6	74.2	68.9	68.4	**72.9**
Se [%]	63.9	62.1	67.1	67.0	66.2	**69.5**
Sp [%]	70.8	71.4	77.1	69.8	69.4	**74.5**

on *OSA_cc*, whereas the EDR combinations only obtain AUCs of 76.3 (Ramp) and 75.1 % (RS slope). For the *OSA_Varon* data set a similar AUC of 75.9 % is obtained using the EDR Ramp, as can be seen in the first column of Table 7.

The classification results on the *OSA_cc* data for segments with clean capacitive sensor data are presented in Table 5. Overall, lower sensitivities can be observed compared to Table 4 in the reduced data set, even for the PSG based classification. Nevertheless, when the combination of ECG with the TB is compared between the PSG ECG and ccECG signals, only a slight drop in AUC of about 1 % is observed. However, further analysis has shown that when using the TB, the respiratory features are dominant in the classifier. When a model was retrained using only the three features extracted from the TB (std, kurtosis and PSD BW resp), still a performance with an AUC, Acc, Se and Sp of, respectively, 80.3, 77.9 57.6 and 86.7 % was obtained.

Substituting the TB by the Ramp EDR signal, causes only a slight drop in AUC of 3.4 % for the PSG ECG signal, whereas, for the ccECG signal this causes a drop of 12.7 %. Using the RS slope EDR estimation, a slightly worse performance is observed with a similar AUC of 67.7 %, but a sensitivity of only 50.3 % for the ccECG data. On the other hand, when comparing the PSG based combination of PSG ECG and TB

Table 5: Apneic segment classification performance on the *OSA_cc* data for segments with clean capacitive sensor data.

| Respiration | TB | | EDR Ramp | | TB | ccBioz | |
ECG	PSG[*]	ccECG	PSG[*]	ccECG	PSG[**]	PSG[**]	ccECG
AUC [%]	81.2	**80.1**	77.8	67.4	81.1	80.4	**79.1**
Acc [%]	77.9	**77.2**	73.7	63.5	83.0	80.7	**80.4**
Se [%]	59.0	**60.7**	54.4	54.4	58.3	58.3	**66.7**
Sp [%]	86.2	**84.3**	82.0	67.4	87.5	84.8	**82.9**

[*] Including only segments with clean ccECG signal.
[**] Including only segments with clean ccECG and ccBioz signal.

against the completely capacitively-coupled combination of ccECG and ccBioz, only slight changes in performance are observed.

4.3. Additional experiments

4.3.1. Comparison of features extracted from PSG and capacitively-coupled measurements Only three of the selected features show a weak to moderate correlation between features computed based on the PSG ECG + EDR Ramp and ccECG + EDR Ramp: the log std RR ($R^2 = 0.31$), WAV_{LF} ($R^2 = 0.12$) and log WAV_{HF} ($R^2 = 0.17$). The other features all have R-squared values below 0.05, meaning these features do not show the same characteristic behavior when extracted from ccECG in comparison to the contact ECG case. As described in (Castro et al., 2018), the most probable reason for this is that these features rely either on the complete ECG morphology or use the ECG for EDR extraction. As a consequence, when only features with an R-squared value larger than 0.1 were considered, all respiratory features were removed, pointing towards an inaccurate EDR extraction from the ccECG. Feature selection on the remaining features resulted in a set of five features, including the three features mentioned above, the log mean of the RR-intervals and the power in the HF band of the RR-intervals.

When a model was trained on the *OSA_ref* data set using these five features, and tested on the ccECG + EDR Ramp signals, a slight increase in performance was observed. An AUC, accuracy, sensitivity and specificity of, respectively, 68.0, 67.5, 48.7 and 75.6 % were obtained. The sensitivity was, however, further decreased. This indicates the need of a good respiratory measure in order to obtain an acceptable sensitivity of the classifier. Better EDR estimates of the ccECG signal, or alternative unobtrusive signals such as ccBioz, are thus needed.

4.3.2. Training a model based on capacitively-coupled data A model was trained based on the ccECG data of the *OSA_cc* set using 10-fold cross validation. The averaged performance measures can be found in Table 6 for both the training and test data. When using an RBF kernel, a test AUC of 62.0 % is obtained, which is worse than the results obtained in Table 5. Nevertheless, the training set performance achieved an AUC of 81.4 %, showing the potential of retraining based on the ccECG. When more ccECG training data would have been available, a better generalization might have been reached. However, with this data set size, the model overfitted to the training set, and high variances are seen in the performance results.

A simpler model might therefore, be more appropriate. LS-SVM models with a linear kernel were tested. A slightly better test AUC of 63.4 % was achieved. In this case, the model did not overfit to the training set, and the training and test performances are similar. It is thus not clear if similar accuracies to the ones obtained using the PSG data are achievable, even with a larger data set.

Table 6: 10-fold cross validated apnea segment classification performance for model trained on the capacitively-coupled data. Mean performance measures ± standard deviation are presented.

	RBF kernel		Linear kernel	
	Training	Test	Training	Test
AUC [%]	81.4 ± 4.6	62.0 ± 7.4	64.6 ± 1.2	63.4 ± 6.9
Acc [%]	74.1 ± 4.5	58.4 ± 5.9	61.2 ± 1.4	58.6 ± 9.2
Se [%]	74.3 ± 3.7	53.8 ± 14.8	64.1 ± 4.1	62.2 ± 17.7
Sp [%]	74.0 ± 5.0	62.9 ± 11.0	60.0 ± 3.6	57.8 ± 15.4

Table 7: Performance comparison to the study by Varon et al. (Varon et al., 2015a)

	Trained on *OSA_ref*	Retrained on *OSA_Varon*	Original study (Varon et al., 2015a)
AUC [%]	75.9	85.2	90.0
Acc [%]	68.7	79.9	75.9
Se [%]	66.1	64.1	78.8
Sp [%]	69.9	86.0	84.6

4.3.3. Retraining the model with OSA_Varon Retraining the classifier using a subset the *OSA_Varon* data set caused an increase in AUC performance of 10 % to 85.2% for the test portion of this data set, as can be seen in Table 7.

5. Discussion

5.1. Selected features

The seven selected features are very comparable to those selected in (Varon et al., 2015a). Only the kurtosis of the respiration and the PSD BW resp feature were not considered by (Varon et al., 2015a), but the serial correlation coefficient was included instead. A lot of different methodologies were tested to estimate the cardiorespiratory interactions, of which the wavelet based subspace projection features were selected in the final set. This might be due to the fact that the wavelet decomposition acts as a filter which smooths out estimation errors in the EDR, while the other cardiorespiratory features take the unfiltered EDR signals as input. This would also explain the relatively high correlation between ECG and ccECG estimates for these features, as seen in Section 4.3.1, compared to the other respiratory features which all obtained R-squared values below 0.1. From these R-squared values, it is clear that the EDR-related features and the QRS_{morph} feature were not accurately estimated from the ccECG signal. This could be expected because these are largely dependent on the QRS morphology, while the ccECG signal can change between different leads from segment to segment and is obtained from

different electrode locations than the PSG ECG. This causes the QRS morphology not to be consistent with that of the PSG ECG. In order to improve the EDR estimate from this ccECG signal, further research is needed related to the effect respiration has on the ccECG signals. The obtained results suggest that the ccBioz signals are a better source to extract respiratory features than the EDR from the ccECG signals. The possibility to have a fully unobtrusive apnea detection solution using ccECG and ccBioz measurements is therefore promising and should be further validated in a larger data set.

5.2. Apnea detection

On the PSG data, our algorithm obtained an ECG based AUC of 76.3 % and an AUC of 80.0 % when the TB was included. This performance might seem rather low compared to studies from the literature in which AUCs of 89-94 % are reported (Mendonca et al., 2018). But most of the studies use the Physionet Apnea-ECG data set. This data set is rather easy to classify, since the control subjects are volunteers instead of patients with sleep related complaints (Penzel et al., 2000). When algorithms trained on the Apnea-ECG data set are tested on clinical data with more complex disease patterns, drops in accuracy of up to 15 % can be seen (Papini et al., 2018). Therefore, only a direct comparison to the algorithm by Varon et al. (Varon et al., 2015a) was included, since this clinical data set, *OSA_Varon*, was available for this study. The results of this comparison using the Ramp EDR can be found in Table 7. Applying our pre-trained model to the *OSA_Varon* data set gave similar performances for the other two data sets. However, the AUC of 75.9 % is much lower than the AUC of 90.0 % in (Varon et al., 2015a). This performance drop might be explained by the difference in apnea scoring rules between *OSA_ref* and *OSA_Varon*, but also by a difference in patient demographics. In Table 1, it can be observed that *OSA_Varon* contains more men, of which a larger percentage has moderate to severe sleep apnea resulting in a larger average AHI and a smaller range of AHI values. When analyzing the different apnea subtypes, it was discovered that in this data set of all segments labeled as apneic, containing apneas or hypopneas, 22 % contained obstructive apneas. These are in general the most severe events, and the easiest to automatically detect. Whereas for the *OSA_ref* and *OSA_cc* data sets this was, respectively, only 10.9 and 11.8 %. When the model was retrained on the *OSA_Varon* data set, the AUC on the *OSA_Varon* test subset increased to 85.2 %. This is still slightly lower than the results in (Varon et al., 2015a), this drop is due to the lowered sensitivity, which is caused by the change in feature set since this one was optimized for *OSA_ref*.

Another difficulty with using a clinical data set is the fact that OSA patients often suffer from cardiovascular (CV) comorbidities (Bradley and Floras, 2009). These comorbidities will distort the RR-interval rhythm and the cardiorespiratory interactions, and might thus distort the extraction of the proposed features. In our study, the presence of CV comorbidities was not considered as an exclusion criterion, since otherwise more

than half of the patients coming into the sleep lab for diagnosis would have been excluded. For the *OSA_ref* set, the presence of CV comorbidities was recorded as explained in (Deviaene et al., 2020). Of those patients, 151 had a CV comorbidity at the time of the PSG, 46 were controls and for the other patients the CV data was incomplete. When the apnea detection performance using the Ramp EDR was compared between them, a slight drop in AUC to 73.1 % was observed for the CV group, compared to an AUC of 75.3 % for the controls. Further investigation into the differences between those two groups also revealed an increased correlation between changes in RR-intervals and the EDR time series, measured using the PRSA parameters, for the CV comorbidity group during normal sleep. This increased correlation was not present when the EDR time series was substituted by the TB. This might indicate that in these patients, the EDR methods do not capture the true cardiorespiratory interactions, but rather just the changes in heart rate, which should be studied further.

In (Varon et al., 2015a), the Ramp method was used to compute the EDR. This method relies on the amplitude of the R-peaks. In the case of ccECG signals, large amplitude shifts can occur. Therefore, the RS slope EDR estimate was proposed as an alternative. From the comparison in (Varon et al., 2020), it was concluded that this was one of the most robust EDR methods. Our results, however, show almost no difference when applied on the PSG ECG, and a slightly better performance for the Ramp method on the ccECG. When the feature selection described in Section 3.2.2 was repeated with the RS slope EDR data instead of the Ramp EDR, the same feature set was obtained.

The fact that the RS slope method is still dependent on the QRS morphology might explain these results, further research into the effect that respiration has on the ccECG is needed. On the other hand, the ccBioz signal could be a good alternative to extracting the EDR from the ccECG which should be further investigated and tested on a larger population.

The results in Table 5 show a decreased sensitivity (for both PSG-based and capacitively-coupled based classification) when only segments containing clean capacitive data are considered. This can be explained by the fact that apneic events which have a clear cardiorespiratory reaction, due to an autonomic arousal, will be more likely to cause (movement) artefacts distorting the capacitive signals. These segments will thus be removed from the analysis due to the lower signal quality and the apneic events left in the data set will have a more subtle cardiorespiratory reaction which is more difficult to detect. This explanation is supported by the observed apnea sub-type division, when considering the complete *OSA_cc* data set, 64.9 % of apneic segments were linked to the less severe hypopnea events which only have a partial cessation of breathing. While for the subset with clean ccECG signals, this percentage increased to 74.9 %. In this study, segments with a low signal quality were discarded, but the periodicity and length of artefacts could also be studied in further research to investigate the link between apneic arousals and certain artefacts, as was proposed in (Huysmans et al., 2019).

The fully unobtrusive apnea classification based on the ccBioz signal combined with

the ccECG has shown good results, comparable to the ones obtained using PSG ECG and TB. These results are, however, based on a very small subset of the data set since quite some ccBioz data was lost due to sensor failure. This subset only contained 263 normal sleep minutes and 48 apneic minutes. Moreover, the ccBioz processing contained a manual artefact detection step. In order to obtain a fully automated method based on the ccBioz signal, an algorithm estimating the signal quality index should be developed. The obtained results using the ccBioz are promising, but should be further validated on a larger data set. New data set collections with the capacitively-coupled sensors are planned. These capacitively-coupled sensors could in the future also be used to assess other medical conditions, next to OSA, during sleep. The recording setup used in this study with the capacitively-coupled sensors embedded in the mattress is still in a prototype stage. In the future, these sensors should be embedded in a flexible mat. Which could be easily installed below the bed cover and transported if necessary.

6. Conclusion

This paper presented an algorithm for the automatic detection of sleep apnea based on (capacitively-coupled) ECG and unobtrusive respiratory measurements. A set of ECG, respiratory and cardiorespiratory parameters were extracted from different combinations of PSG or capacitively coupled ECG signals with EDR measures, thoracic belt or ccBioz signals. Using the PSG data, slightly better performances were obtained when using the thoracic belt instead of the EDR. When tested only on the unobtrusive ccECG data, a clear drop in performance was observed, which was linked to difficulties extracting the EDR from the ccECG due to morphological changes within the signal. When the ccECG signal was instead combined with the TB or ccBioz, similar performances as for the PSG data set were obtained. Especially the results on the combination of ccECG with ccBioz are very promising, as this paper is the first to test this fully unobtrusive solution. These results, should however, be further validated on a larger data collection of ccBioz signals.

Acknowledgements

References

Adult Obstructive Sleep Apnea Task Force of the AASM (2009). Clinical guideline for

the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med*, 5(3):263–276.

Bauer, A., Kantelhardt, J. W., Bunde, A., Barthel, P., Schneider, R., Malik, M., and Schmidt, G. (2006). Phase-rectified signal averaging detects quasi-periodicities in non-stationary data. *Physica A*, 364:423–434.

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S., Morrell, M. J., Nunez, C. M., Patel, S. R., Penzel, T., Pépin, J.-L., et al. (2019). Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir. Med.*, 7(8):687–698.

Berry et al, R. B. (2012). Rules for scoring respiratory events in sleep: update of the 2007 aasm manual for the scoring of sleep and associated events. *J Clin Sleep Med*, 8(5):597–619.

Bradley, T. D. and Floras, J. S. (2009). Obstructive sleep apnoea and its cardiovascular consequences. *Lancet*, 373(9657):82–93.

Castro, I., Varon, C., Moeyersons, J., Gomez, A. V., Morales, J., Deviaene, M., Torfs, T., Van Huffel, S., Puers, R., and Van Hoof, C. (2019a). Data quality assessment of capacitively-coupled ecg signals. In *CinC 2019*, pages 1–4. IEEE.

Castro, I. D., Patel, A., Torfs, T., Puers, R., and Van Hoof, C. (2019b). Capacitive multi-electrode array with real-time electrode selection for unobtrusive ecg & bioz monitoring. In *EMBC 2019*, pages 5621–5624. IEEE.

Castro, I. D., Varon, C., Torfs, T., Van Huffel, S., Puers, R., and Van Hoof, C. (2018). Evaluation of a multichannel non-contact ecg system and signal quality algorithms for sleep apnea detection and monitoring. *Sensors*, 18(2):577.

De Brabanter, J., De Moor, B., Suykens, J. A., Van Gestel, T., and Vandewalle, J. P. (2002). *Least squares support vector machines*. World Scientific.

De Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P., and O'Malley, M. (2003). Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE TBME*, 50(6):686–696.

Deviaene, M., Borzée, P., Van Gilst, M., Van Dijk, J., Overeem, S., Buyse, B., Testelmans, D., Van Huffel, S., and Varon, C. (2020). Multilevel interval coded scoring to assess the cardiovascular status of sleep apnea patients using oxygen saturation markers. *IEEE TBME*, 67(10):2839–2848.

Deviaene, M., Testelmans, D., Borzée, P., Buyse, B., Van Huffel, S., and Varon, C. (2019). Feature selection algorithm based on random forest applied to sleep apnea detection. In *EMBC 2019*, pages 2580–2583. IEEE.

Deviaene, M., Testelmans, D., Buyse, B., Borzée, P., Van Huffel, S., and Varon, C. (2018). Automatic screening of sleep apnea patients based on the spo 2 signal. *IEEE JBHI*, 23(2):607–617.

Faust, O., Acharya, U. R., Ng, E., and Fujita, H. (2016). A review of ecg-based diagnosis support systems for obstructive sleep apnea. *J Mech Med Biol*, 16(01):1640004.

Guilleminault, C., Winkle, R., Connolly, S., Melvin, K., and Tilkian, A. (1984). Cyclical variation of the heart rate in sleep apnoea syndrome: mechanisms, and usefulness of 24 h electrocardiography as a screening technique. *Lancet*, 323(8369):126–131.

Huysmans, D., Borzée, P., Testelmans, D., Buyse, B., Willemen, T., Van Huffel, S., and Varon, C. (2019). Evaluation of a commercial ballistocardiography sensor for sleep apnea screening and sleep monitoring. *Sensors*, 19(9):2133.

Lázaro, J., Alcaine, A., Romero, D., Gil, E., Laguna, P., Pueyo, E., and Bailón, R. (2014). Electrocardiogram derived respiratory rate from qrs slopes and r-wave angle. *Ann Biomed Eng*, 42(10):2072–2083.

Malik, M. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the european society of cardiology and the north american society for pacing and electrophysiology. *Ann Noninvas Electro*, 1(2):151–181.

Mendonca, F., Mostafa, S. S., Ravelo-García, A. G., Morgado-Dias, F., and Penzel, T. (2018). A review of obstructive sleep apnea detection approaches. *IEEE JBHI*, 23(2):825–837.

Moeyersons, J., Smets, E., Morales, J., Villa, A., De Raedt, W., Testelmans, D., Buyse, B., Van Hoof, C., Willems, R., Van Huffel, S., et al. (2019). Artefact detection and quality assessment of ambulatory ecg signals. *Comput Meth Prog Bio*, 182:105050.

Morales, J. F., Deviaene, M., Milagro, J., Testelmans, D., Buyse, B., Willems, R., Orini, M., Van Huffel, S., Bailón, R., and Varon, C. (2019). Evaluation of methods to characterize the change of the respiratory sinus arrhythmia with age in sleep apnea patients. In *EMBC 2019*, pages 1588–1591. IEEE.

Papini, G. B., Fonseca, P., Margarito, J., van Gilst, M. M., Overeem, S., Bergmans, J. W., and Vullings, R. (2018). On the generalizability of ecg-based obstructive sleep apnea monitoring: merits and limitations of the apnea-ecg database. In *EMBC 2018*, pages 6022–6025. IEEE.

Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L., and Peter, J. H. (2000). The apnea-ecg database. In *CinC 2000*, pages 255–258. IEEE.

Quan, S., Gillin, J. C., Littner, M., and Shepard, J. (1999). Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. editorials. *Sleep*, 22(5):662–689.

Spinelli, E. and Haberman, M. (2010). Insulating electrodes: a review on biopotential front ends for dielectric skin–electrode interfaces. *Physiol meas*, 31(10):S183.

Uddin, M., Chow, C., and Su, S. (2018). Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review. *Physiol meas*, 39(3):03TR01.

Varon, C., Caicedo, A., Testelmans, D., Buyse, B., and Van Huffel, S. (2015a). A novel algorithm for the automatic detection of sleep apnea from single-lead ecg. *IEEE TBME*, 62(9):2269–2278.

Varon, C., Jansen, K., Lagae, L., and Van Huffel, S. (2015b). Can ecg monitoring identify seizures? *J Electrocardiol*, 48(6):1069–1074.

Varon, C., Morales, J., Lázaro, J., Orini, M., Deviaene, M., Kontaxis, S., Testelmans, D., Buyse, B., Borzée, P., Sörnmo, L., et al. (2020). A comparative study of ecg-derived respiration in ambulatory monitoring using the single-lead ecg. *Scientific Reports*, 10(1):1–14.