Accelerometry-Derived Respiratory Index estimating Apnea-Hypopnea Index for Sleep Apnea Screening

Aurélien Bricout, Julie Fontecave-Jallon, Jean Louis Pepin, P.Y. Gumery

▶ **To cite this version:**

Aurélien Bricout, Julie Fontecave-Jallon, Jean Louis Pepin, P.Y. Gumery. Accelerometry-Derived Respiratory Index estimating Apnea-Hypopnea Index for Sleep Apnea Screening. Computer Methods and Programs in Biomedicine, 2021, 207, pp.106-209. 10.1016/j.cmpb.2021.106209 . hal-03680408

HAL Id: hal-03680408
https://hal.science/hal-03680408v1

Submitted on 13 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0169260721002832
Manuscript_792e82eade9a5c0f1d15338ad5e83f19

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Accelerometry-Derived Respiratory Index estimating Apnea-Hypopnea Index for Sleep Apnea Screening

Aurélien Bricout[1], Julie Fontecave-Jallon[1], Jean-Louis Pépin[2] and Pierre-Yves Guméry[1]

[1]*Grenoble Alpes University, CNRS, CHU Grenoble Alpes, Grenoble INP, TIMC, Grenoble, France (phone: + 33 04 56 52 00 62 ; e-mail: aurelien.bricout@gmail.com or pierre-yves.gumery@univ-grenoble-alpes.fr)*

[2]*HP2 Laboratory, INSERM U1042, Grenoble Alpes University, Grenoble, France; e-mail@e-mail.com*

ARTICLE INFO

Article history:

Received
Revised

Keywords:

Polysomnography
Screening
Sleep Apnea Syndrome
Machine Learning
Accelerometry
Respiration

ABSTRACT

Background and Objective: Sleep Apnea Syndrome (SAS) is a multimorbid chronic disease with individual and societal deleterious consequences. Polysomnography (PSG) is the multi-parametric reference diagnostic tool that allows a manual quantification of the apnea-hypopnea index (AHI) to assess SAS severity. The burden of SAS is affecting nearly one billion people worldwide explaining that SAS remains largely under-diagnosed and undertreated. The development of an easy to use and automatic solution for early detection and screening of SAS is highly desirable.
Methods: We proposed an Accelerometry-Derived Respiratory index (ADR) solution based on a dual accelerometry system for airflow estimation included in a machine learning process. It calculated the AHI thanks to a RUSBoosted Tree model and used physiological and explanatory specifically developed features. The performances of this method were evaluated against a configuration using gold-standard PSG signals on a database of 28 subjects.
Results: The AHI estimation accuracy, specificity and sensitivity of the ADR index were 89%, 100% and 80% respectively. The added value of the specifically developed features was also demonstrated.
Conclusion: Overnight physiological monitoring with the proposed ADR solution using a machine learning approach provided a clinically relevant estimate of AHI for SAS screening. The physiological component of the solution has a real interest for improving performance and facilitating physician's adhesion to an automatic AHI estimation.

1. Introduction

Sleep apnea is one of the most frequent chronic disease affecting one billion people worldwide and associated with cardiometabolic comorbidities and an increased risk of mortality [1][2]. It is characterized by the occurrence of frequent and abnormally distributed episodes of stoppage (apnea) or significant reduction (hypopnea) of respiratory airflow lasting at least ten seconds [3].

Complete (apneas) or partial (hypopneas) airflow reduction occur during sleep ended by micro-arousals producing sleep fragmentation. The nature of the events can be obstructive (pharyngeal collapses) and is then characterized by the preservation of thoracic and abdominal movements. Airflow cessation or reduction can also be of central origin, characterized by a reduction of respiratory centers drive and is associated with the absence of thoracic and abdominal movements [4]. Hypopnea is defined by a partial airflow reduction (30% or 50% depending on the scoring rules) of more than 10 seconds, followed by a micro-arousal or an oxygen desaturation of at least 3%.

© 2021 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

The severity of SAS is assessed by the Apnea-Hypopnea Index (AHI), which is the number of abnormal respiratory events per hour of sleep or recording, according to the following criteria [5]:

- AHI less than 5: absence of the syndrome
- AHI between 5 and 15: mild syndrome
- AHI between 15 and 30: moderate to severe syndrome
- AHI greater than 30: severe syndrome

Polysomnography (PSG) is the gold standard method used in sleep laboratories or at home, for the reference diagnosis of sleep disorders, including SAS. It is a multi-parametric assessment, which simultaneously records data from several channels such as electroencephalography (EEG), electrocardiography (ECG), airflow, oxygen saturation and respiratory movements measured by Respiratory Inductance Plethysmography (RIP) [6]. For the diagnosis of SAS, PSG recordings are manually scored to quantify the Apnea-Hyponea Index (AHI). This is time consuming and requires human expertise with a preferential use of airflow measured by a nasal cannula, thoracic and abdominal respiratory movements (RIP) and oxygen saturation to score and characterize apneas and hypopneas.

Considering the high prevalence and burden of sleep apnea [2], the current diagnosis tools are inappropriate to face the size of the problem. In laboratory, PSG is a complex and costly diagnostic process, with long waiting lists with inequalities in access to care [7]. These long waiting lists are due to the congestion of sleep centers and the long and specialized time required to score polysomnographic recordings.

In this context, in recent years, several studies have proposed the use of machine learning or deep learning methods for the automatic detection and scoring of sleep apnea ([8], [9], [10], [11], [12] and [13]). The study [13], for example, proposed the use of a neural network model using oxygen saturation features for the detection of positive SAS patients and showed an accuracy of 93.3%. However, a high proportion of these studies used the same database for feeding their models (Physionet Apnea-ECG Database [14]). In addition, several devices for sleep apnea screening have also been proposed in the recent years in order to facilitate the screening process such as the WatchPAT™ developed by Itamar Medical LTD [15], the ApneaLink™, developed by ResMed [16] or the HealthPatch™ developed by VitalConnect [17]. The validation of HealthPatch™ proposed by Nandakumar Selvaraj and Ravi Narasimhan in 2015 showed promising results of classification of subjects with an AHI > 15 with an accuracy of 89.4% using an integrated system combining one accelerometer and an ECG lead in a specific hardware solution supported by a machine learning approach.

In the present study, we proposed the use of accelerometry-derived respiratory index (ADR) system composed of two accelerometers patched on the subject's chest. The use of accelerometry in sleep monitoring had been suggested, for example in [18], [19] or [20]. This kind of system could help to reduce the sensors congestion and facilitate the screening process and the conditions of use. However, none of these studies gave access to an airflow estimation which is of major interest in SAS detection.

The solution proposed in this study included an airflow estimation algorithm based on the thoracic and abdominal efforts measured by the two accelerometers. This algorithm has already been validated during sleep period without abnormal events in our previous study [21] and results demonstrated the good feasibility of the use of an adaptive ADR method for respiration monitoring in this context.

The present paper is then the logical continuation of [21] by proposing a pathophysiological validation of this ADR technology in patients referred for SAS suspicion. The aim is to use a machine learning approach for the validation of the proposed device by evaluating its ability to detect abnormal events using specific and explanatory physiological features. These features have first been extracted and tested on an airflow signal from the nasal cannula, the gold-standard sensor for airflow monitoring in PSG. An automatic AHI estimation based on a classification model has then been implemented. Several inputs from PSG were used in addition to the ADR system in order to validate this approach in a polysomnographic context and compare ADR to the reference nasal cannula.

Materials and Methods

2.1. Sleep Study Data Acquisition

A sleep study protocol was considered and included 28and untreated volunteer SAS subjects at the sleep laboratory of Grenoble University Hospital for an overnight polysomnography. They provided written information consent and the study was approved by the relevant ethics committee (CHU Grenoble Alpes). The AHI had a range of 1.4 – 74.8 and a mean ± standard-

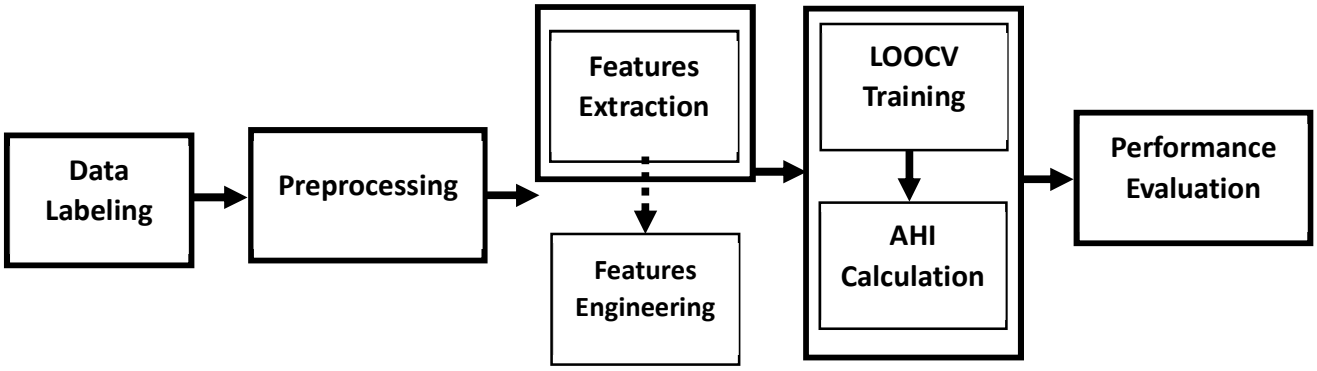

Figure 1: Data processing flow chart

deviation of 14.6±12.2. The BMI had a mean value ± standard-deviation of 25.5±4.8 kg/m².

Subjects were equipped with classical complete PSG system and an ADR accelerometry system. The Deltamed 32-channel PSG system (EEG Brainbox 1042, Natus, Pleasanton, California, USA) was used to collect the standard PSG data at a sampling frequency of 256Hz. Among PSG signals, thorax and abdomen cross sectional area changes (noted RIP_{THO} and RIP_{ABD}) were recorded thanks to Respiratory Inductance Plethysmography (RIP). Nasal airflow (noted $AIRFLOW_{PSG}$) was measured by the nasal cannula, cardiac activity (noted ECG) was measured by an electrocardiogram, and oxygen saturation (noted SpO_2) was measured by a finger pulse oximeter. The ADR system consisted in two accelerometers (STMicroelectronics, LIS344ALH, 3 axes, Analog), placed on the thorax and on the abdomen of the subject. Accelerometers data noted $[X,Y,Z]_{THO}$ and $[X,Y,Z]_{ABD}$ were synchronously collected with the PSG data at 256Hz (imposed by the Natus system).

The data processing included then the succession of several steps as illustrated in Fig 1 and described in the sections below. All developments were implemented in Matlab (Mathworks, R2018b).

2.2. *Data Labeling*

After data acquisition, a first step of formatting and data labeling was applied. Annotations scored by a unique sleep expert for each recording allowed the extraction of the start time, the end time and the type of every abnormal events (obstructive sleep apneas and hypopneas, mixed and central sleep apneas).

Overnight recordings were segmented into different epoch durations (noted EpochDuration) of 60, 150 or 300 seconds. Every epoch was then labeled depending of the absence (negative) or presence (positive) of at least one abnormal event inside the window or a ubiquitous event with at least 50% of its duration inside the considered epoch, as illustrated in Fig 2. The value of 60 seconds is based on the important amount of studies using 60s segmentation epoch, in particular those using Physionet Apnea-ECG Database [14]. The value of 150 seconds is based on the validation study of the HealthPatch™ [17]. The value of 300 seconds is based on a physiological assumption that Heart-Rate Variability (HRV) features are relevant when extracted from at least 5 minutes' window [22]. Furthermore, sleep respiratory events are annotated by experts using 5 minutes' window

2.3. *Preprocessing*

Each signal of interest (called input in the following) was preprocessed to make them suitable for machine

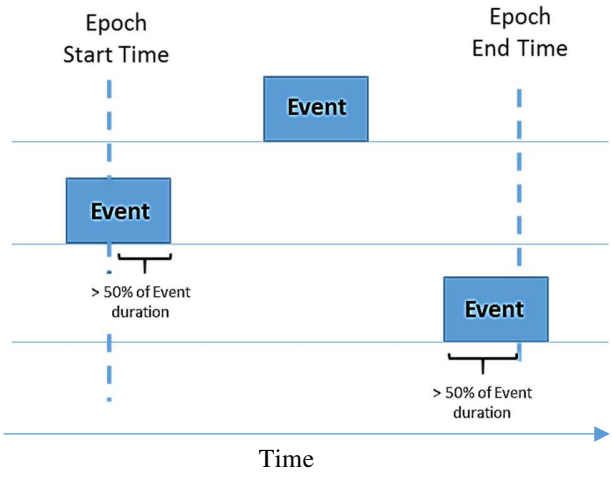

Figure 2: Positive epoch labeling method – 3 examples of positive epochs

learning process. $AIRFLOW_{PSG}$, RIP_{THO} and RIP_{ABD}, SpO_2 and Tri-axis accelerometer $[X,Y,Z]_{THO}$ and $[X,Y,Z]_{ABD}$ data were first undersampled at 8Hz. Respiratory signals and accelerometer data were then band-pass filtered between 0.05Hz and 0.8Hz in order to extract only the respiratory component and to remove artefacts and physiological noises. Thoracic and abdominal efforts reconstruction (ADR_{THO} and ADR_{ABD} respectively) from accelerometer data were calculated according to (1) and (2), as described in detail in [21]. A Principal Component Analysis (PCA) of $[X,Y,Z]_{THO}$ and $[X,Y,Z]_{ABD}$ was computed to calculate the column eigenvector $[\mu_{THO}\ \beta_{THO}\ \gamma_{THO}]$ (resp. $[\mu_{ABD}\ \beta_{ABD}\ \gamma_{ABD}]$) that explains the majority of the variance on each compartment.

$$ADR_{THO} = \mu_{THO}.X_{THO} + \beta_{THO}.Y_{THO} + \gamma_{THO}.Z_{THO} \quad (1)$$
$$ADR_{ABD} = \mu_{ABD}.X_{ABD} + \beta_{ABD}.Y_{ABD} + \gamma_{ABD}.Z_{ABD} \quad (2)$$

A volume estimation V_{ADR} (resp. V_{RIP}) was estimated using a linear combination of thoracic and abdominal efforts (resp. thoracic and abdominal cross-sectional variation changes) such as described in (3) and (4) [23].

$$V_{ADR} = \tau.ADR_{THO} + \alpha.ADR_{ABD} \quad (3)$$
$$V_{RIP} = \tau.RIP_{THO} + \alpha.RIP_{ABD} \quad (4)$$

α and τ were set to 2 and 1 such as proposed in [24]. $AIRFLOW_{ADR}$ and $AIRFLOW_{RIP}$ were then computed as the derivate of V_{ADR} and V_{RIP} respectively.

On the other hand, ECG was band-pass filtered between 5Hz and 30Hz to remove artefacts and physiological noises. Two inputs were extracted from ECG signal. The first was the 4Hz interpolated signal of RR intervals (noted RRI) and the second was the 4Hz interpolated signal of QRS amplitude (noted QRS_{AMP}) such as recommended in [22] and described in [25]. QRS complexes were detected using the Pan and Tompkins algorithm [26] [27].

After preprocessing, the inputs available, proposed in the model and used for features engineering, were $AIRFLOW_{PSG}$ (resp. $AIRFLOW_{ADR}$, $AIRFLOW_{RIP}$), SpO_2, RRI and QRS_{AMP}.

2.4. Features extraction and Features engineering

Specific features were extracted from each input in every epoch. Table 1, Table 2, Table 3 and Table 4 enumerate the features extracted from respiratory inputs [17] ($AIRFLOW_{PSG}$, $AIRFLOW_{ADR}$, and $AIRFLOW_{RIP}$), from RRI [17], from QRS_{AMP} [17] and from SpO_2 [28] respectively. Features were then normalized for each subject.

Table 1. Features extracted from AIRFLOW inputs per epoch [17]. Features with * are new propositions explained in detail in the text

Feature Name	Feature Description
Median	Median of epoch input samples
Std	Standard-deviation of epoch input samples
CoeffVar	Coefficient variation of epoch input samples
MeanAD	Mean absolute deviation of epoch input samples
MedianAD	Median absolute deviation of epoch input samples
Kurtosis	Kurtosis of epoch input samples
Iqr	Interquartile range of epoch input samples
DispMetric	Dispersion metric of epoch input samples
VLF	Very Low Frequency power below 0.04Hz
LF	Low Frequency power between 0.04 - 0.15Hz range
HF	High Frequency power between 0.15 - 0.40Hz range
LF/HF	Low Frequency / High Frequency ratio
SpecKurtosis	Spectral Kurtosis
RR	Respiratory rate (Number of respiratory cycles / minute)
*NPRate**	Rate of Non-Periodic 20s-windows rate during the epoch.
*ERate**	Rate of 20s-windows with low energy during the epoch
*VarDispMetric**	Variance of the dispersion metric in every 20s-window during the epoch

Table 2. Features extracted from RRI [17]

Feature Name	Feature Description
HR	Heart rate (Number of cardiac cycles / minute)
SDSD	Standard deviation of successive differences
SDNN	Standard deviation of NN intervals
RMSSD	Root mean square of successive differences
pNN50	Probability of intervals greater or smaller than 50ms
TRI	Triangular index from the interval histogram
TINN	Triangular Interpolation of NN intervals
AppEn	Approximate Entropy
pLF	Percentage of Low Frequency power
PHF	Percentage of High Frequency power
VLF	Very Low Frequency power below 0.04Hz
LF	Low Frequency power between 0.04 - 0.15Hz range
HF	High Frequency power between 0.15 - 0.40Hz range
LF/HF	Low Frequency / High Frequency ratio

SD1	Pointcaré plot standard deviation perpendicular the line of identity (ms)
SD2	Pointcaré plot standard deviation along the line of identity (ms)
SD1SD2	Ratio of SD1 and SD2
Median	Median of epoch input samples
Std	Standard-deviation of epoch input samples
CoeffVar	Coefficient variation of epoch input samples
MeanAD	Mean absolute deviation of epoch input samples
MedianAD	Median absolute deviation of epoch input samples
Kurtosis	Kurtosis of epoch input samples
Iqr	Interquartile range of epoch input samples
DispMetric	Dispersion metric of epoch input samples

Table 3. Features extracted from QRS [17]

Feature Name	Feature Description
Median	Median of epoch input samples
Std	Standard-deviation of epoch input samples
CoeffVar	Coefficient variation of epoch input samples
MeanAD	Mean absolute deviation of epoch input samples
MedianAD	Median absolute deviation of epoch input samples
Kurtosis	Kurtosis of epoch input samples
Iqr	Interquartile range of epoch input samples
DispMetric	Dispersion metric of epoch input samples
AppEn	Approximate Entropy
SampEn	Sample Entropy
VLF	Very Low Frequency power below 0.04Hz
LF	Low Frequency power between 0.04 - 0.15Hz range
HF	High Frequency power between 0.15 - 0.40Hz range
LF/HF	Low Frequency / High Frequency ratio
SpecKurtosis	Spectral Kurtosis

Table 4. Features extracted from SpO_2 inputs [28]

Feature Name	Feature Description
Min	Minimum of epoch SpO_2 samples
Mean	Mean value of epoch SpO_2 samples
Variance	Variance of epoch SpO_2 samples
NumZC	Zero-Crossing Rate using *Mean* as baseline
Slope	Slope of the regression line fitted for epoch SpO_2 samples
AbsSlope	Absolute value of *Slope*
Delta Index	Delta Index
TSA70, TSA80, TSA85, TSA90, TSA95	Accumulative time that SpO_2 level stays below 70, 80, 85, 90, 95
ODIS2, ODIS3, ODIS4, ODIS5	The total number of SpO_2 samples that fall at least 2, 3, 4, 5 below the *Mean* as Baseline
AppEn	Approximate entropy
LempZC	Lempel-Ziv complexity
CTM25, CTM50, CTM75, CTM100	Central Tendency Measure
ODIxy, $x \in \{2, 3, 4, 5\}$ $y \in \{1, 3, 5\}$	The number of occurrences that SpO_2 level declines at least x below the baseline and lasts at least y seconds

Neighborhood Components Analysis (NCA) [30] was implemented using a stochastic gradient descent to calculate the weight of every features used for the classification.

As shown in Table 1, three respiratory features (*NPRate*, *ERate* and *VarDispMetric*) were specifically developed in order to increase the physiological meaning of the model and improve the sensitivity and specificity of the detection.

Since ventilation is defined as a pseudo-periodic phenomenon, the detection of abnormal events could be therefore based on the evaluation of the non-periodicity of the signal. For that purpose, the respiratory signals were segmented into window of 20s with an overlap of 90%. In each window, a Normalized Auto-Correlation (NAC) was calculated such as described in [29] for acoustic signals. The amplitude of the NAC lag zero is considered as reference. The first local maximum of the NAC in positive lags side in the window was detected and compared to the reference. If it was higher than 80% of the lag zero amplitude, the window was considered as periodic such as illustrated in Fig 3.c), for a negative 20s-window (without abnormal events). On the contrary, if lower than 80%, the window was considered as non-periodic, such as illustrated in Fig 3.f) for a positive 20s-window. *NPRate* was therefore defined as the rate of non-periodic windows during the epoch. The *NPRate* for the negative epoch in Fig 3.a) was 0% and for the positive epoch in Fig 3.d) was 62.5%. Indeed, in positive epoch, periodic windows could be found during ventilatory recovery after abnormal events.

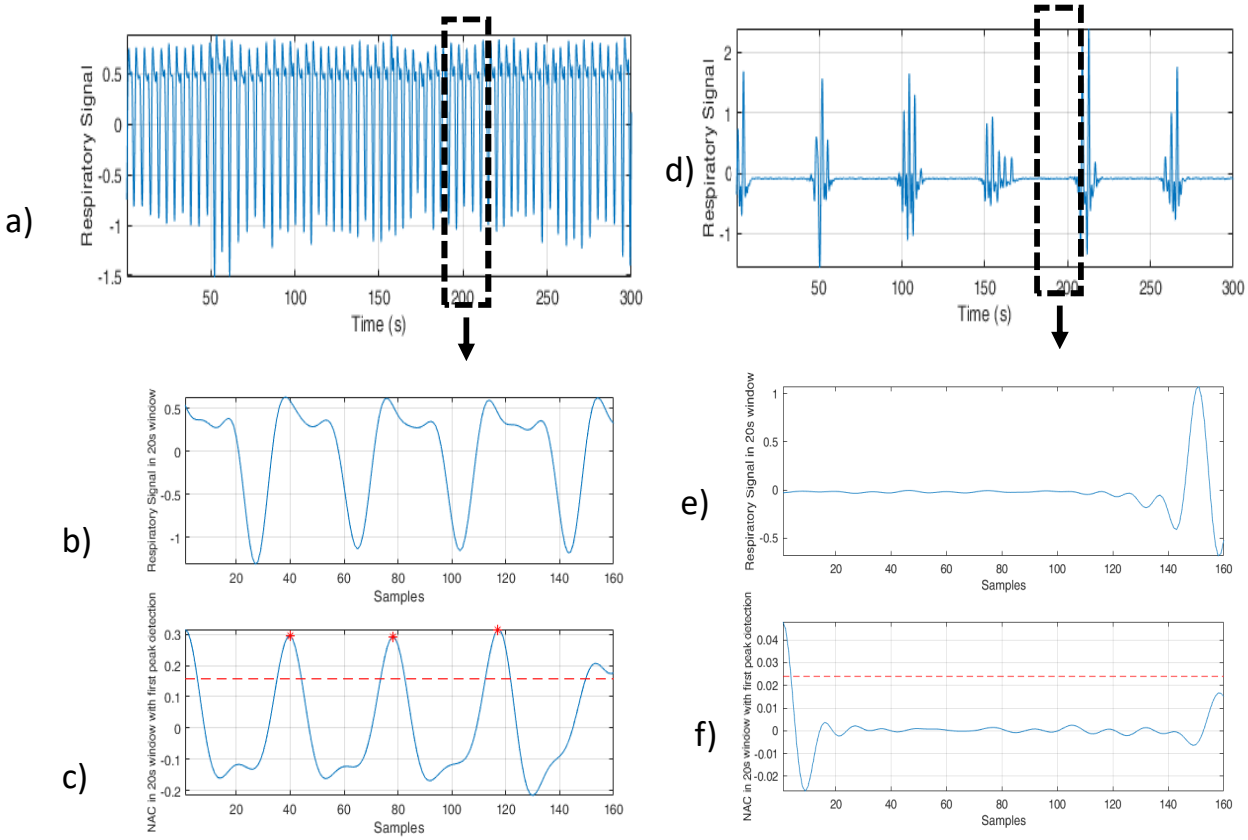

Figure 3: Illustration of the Non-Periodic Window Rate feature (*NPRate)*:

a) Airflow of a negative epoch (*NPRate = 0%)* with an example in a periodic 20s-window b) where the NAC with peak detection c) was performed.

d) Airflow of a positive epoch (*NPRate = 62.5%)* with an example in a non-periodic 20s-window e) where the NAC f) was performed and the peak detection failed.

As mentioned in the introduction, abnormal events are defined as an airflow reduction that leads to a local diminution of the energy contained in the signal. In each 20sec-window, the energy of the signal (noted *E*) was calculated as the first sample of the Normalized Auto-Correlation signal as already mentioned above. If the energy is lower than an empiric threshold (0.2 such as illustrated in Fig 4), the window was considered as abnormal. *ERate* was therefore defined as the rate of window with low energy variance during the epoch. Fig 4.b) shows the *E* calculated in every 20s-window during negative epoch, resulting in an *ERate* of 0%. On the contrary, Fig 4.d) shows the *E* calculated in every 20s-window during positive epoch, resulting in an *ERate* of 81%.

VarDispMetric was the variance of the dispersion metric (*DispMetric*) calculated in every 20s-window during the epoch such as illustrated in Fig 5. Fig 5.b) shows the *DispMetric* calculated in every 20s-windows in a negative epoch, resulting in a relatively low *VarDispMetric* of 0.008. On the contrary, Fig 5.d) shows the *DispMetric* calculated in every 20s-windows in a negative epoch, resulting in a relatively high *VarDispMetric* of 0.18.

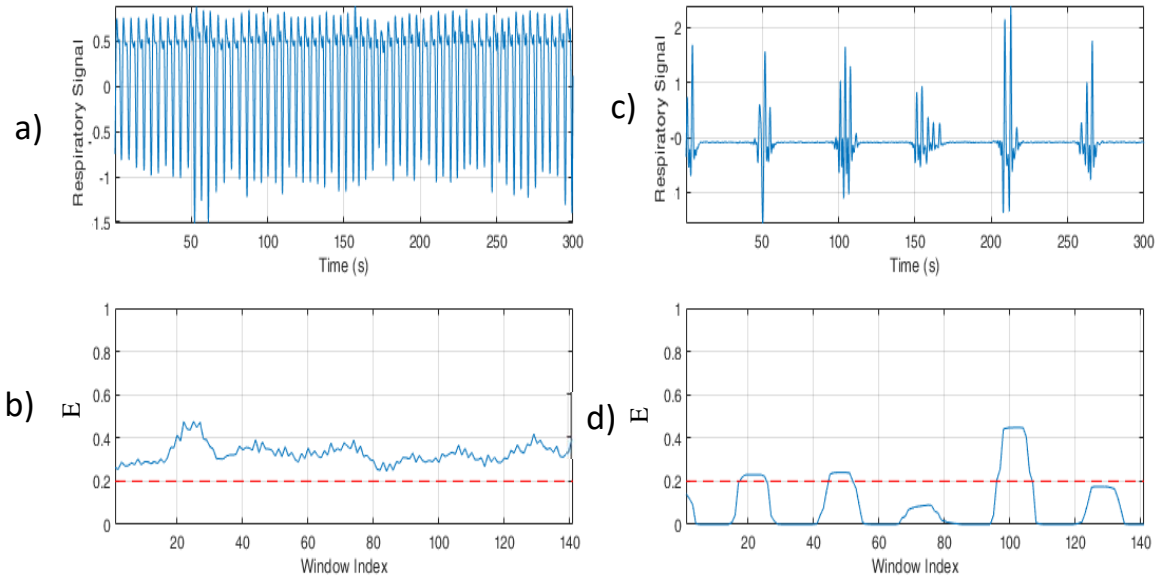

Figure 4: Illustration of the low energy window rate feature (*ERate)* for:

a) Airflow of a negative epoch in which the energy of each 20s-window is relatively high (i.e. higher than a 0.2 threshold) b), resulting in an *ERate = 0%*.

c) Airflow of a positive epoch in which the energy of each 20s-window is relatively low (i.e. lower than a 0.2 threshold) d), resulting in an *ERate = 81%*.

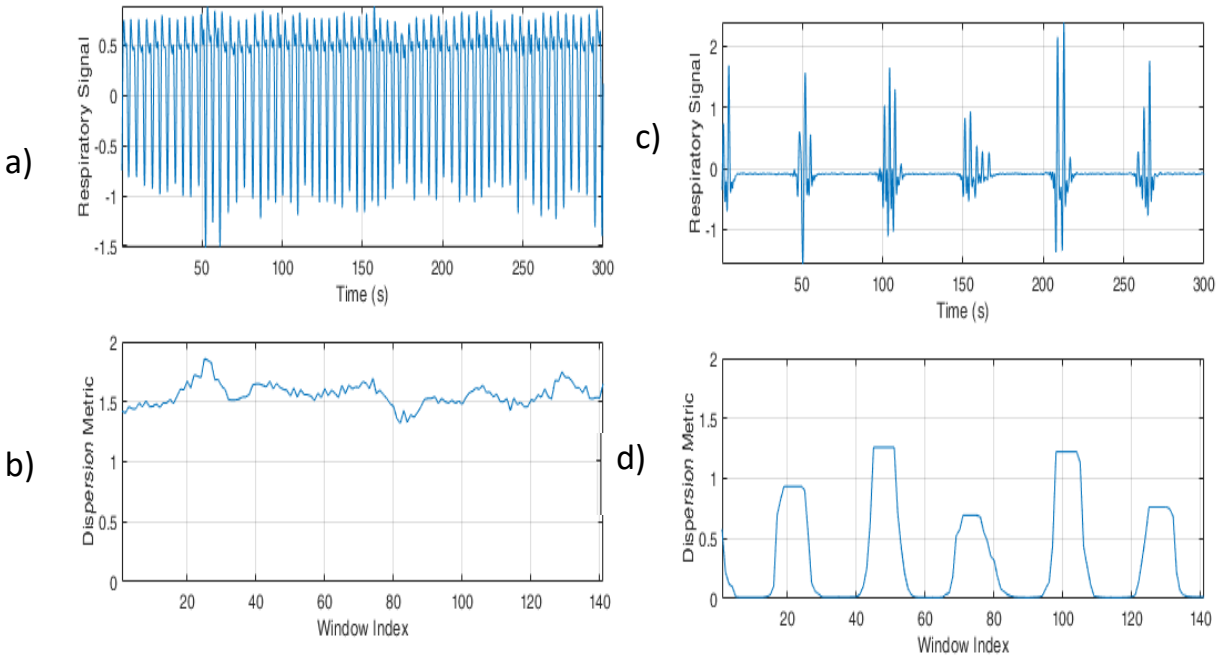

Figure 5: Illustration of the variance of the dispersion metric in every window feature (*VarDispMetric)* for:

a) Airflow of a negative epoch in which the variance of the dispersion metric measured in each 20s-window is relatively low (*VarDispMetric = 0.008)* b).

c) Airflow of a positive epoch in which the variance of the dispersion metric measured in each 20s-window is relatively high (*VarDispMetric = 0.18)* d).

2.6. Classification and AHI Calulcation

Detection of positive epochs using LOOCV

Training and classification were performed using a Leave-One-Out Cross Validation (LOOCV) approach. On one iteration, twenty-seven subjects were considered for training and the model was tested on the remaineing subject to detect positive epochs i.e. epoch with at least one abnormal event. Positive Epoch per Hour (PEH_{EST}) were then calculated for each subject as the number of positive epochs out of the time of sleep provided by sleep expert scoring. Several classifiers were implemented and tested (such as Support Vectors Machine using Linear, Polynomial and Gaussian approach or Linear Discriminant Analysis) but RUSBoosted Trees (Random Undersamplig) [31] was the classifier with the best approach in terms of class imbalance management, computing time and classification performance.

AHI Estimation

After the calculation of PEH_{EST}, a regression method was used such as proposed in [17]. Using the reference number of positive epoch per hour according to expert annotations from PSG recordings (PEH_{PSG}) and the reference AHI provided by the experts (AHI_{PSG}), a regression model was identified for each epoch size (60s, 150s and 300s) such as $AHI_{PSG} = f_{EpochSize}(PEH_{PSG})$, where $f_{EpochSize}$ is the regression rule as illustrated in Fig 6.

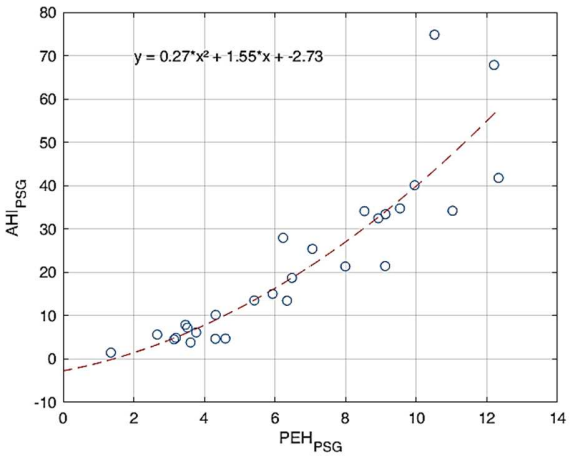

Figure 6: Regression rule implementation for EpochSize = 300s

The regression rule $f_{EpochSize}$ was then applied to PEH_{EST} in order to calculate AHI_{EST}.

2.7. Performance evaluation

Several performance indicators were proposed in order to evaluate the detection of positive epochs and therefore the screening of sleep apnea syndrome based on a AHI higher than 15. This threshold was chosen to separate subjects with no syndrome or a mild syndrome (13 subjects over the 28 subjects of the database) from subjects with a moderate or severe syndrome (15 subjects over the 28 subjects of the database).

The first indicator was the mean F1-Score ± standard deviation for each subject for the performance of the detection of positive epochs, such as defined in (5) where the Recall is defined in (8) and Precision in (10). The higher the F1-Score, the better the performance.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

The second indicator was the mean Bias ± standard deviation between the estimated AHI and AHI from PSG for each subject, such as defined in (6). The closer the bias is to 0, the higher the performance.

$$Biais = AHI_{PSG} - AHI_{EST} \quad (6)$$

The following indicators were the accuracy, sensitivity and specificity, extracted from the screening confusion matrix, such as illustrated in Fig 7. These three indicators also reflected the performance of the estimation of AHI_{EST} in comparison to AHI_{PSG}.

	$AHI_{EST} < 15$	$AHI_{EST} > 15$
$AHI_{PSG} < 15$	True Negative (TN)	False Positive (FP)
$AHI_{PSG} > 15$	False Negative (FN)	True Positive (TP)

Figure 7: Screening confusion matrix

Accuracy (Acc), Sensitivity or Recall (Sen) and Specificity (Spe) were calculated such as defined in (7), (8) and (9) respectively. The Precision was calculated for the measure of the F1-Score such as defined in (10).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Sen = \frac{TP}{TP + FN} \ (= Recall) \quad (8)$$

$$Spe = \frac{TN}{TN + FP} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Performance was first evaluated using features from PSG inputs, *i.e.* $AIRFLOW_{PSG}$, SpO_2, RRI and QRS_{AMP} (called $CONF_{PSG}$) and with the various sizes of epochs.

Then, the same configuration was considered but $AIRFLOW_{PSG}$ from nasal cannula was replaced by $AIRFLOW_{ADR}$ (resp. $AIRFLOW_{RIP}$), called $CONF_{ADR}$ ($CONF_{RIP}$ respectively) and the performances of these 3 configurations were compared to validate the use of the ADR index in a polysomnography context for a screening approach. Results were also compared to previous works.

3. Results

Several classification parameters such as the number of learning cycle, learning rate and misclassification cost were optimized for the $CONF_{PSG}$ in order to get the best model performance and these parameters were then used for $CONF_{RIP}$ and $CONF_{ADR}$.

3.1. Determination of epoch size

Table 6 presents the results and performance indicators for $CONF_{PSG}$, $CONF_{ADR}$ and $CONF_{RIP}$ depending on EpochSize. Overall, results showed that 300s was the best epoch size compared to 60s and 150s achieving an accuracy for the screening classification of 100% for $CONF_{PSG}$ and 89% for $CONF_{RIP}$ and $CONF_{ADR}$. Bias analysis showed that 300s was also the best epoch choice, lower than approximately five events per hour compare to 60s for the $CONF_{PSG}$ and $CONF_{RIP}$ and two events per hour for $CONF_{ADR}$. For the three configurations, with 300s epoch size, F1-Score was approximately 8 to 10% higher than 60s epoch size and 2 to 4% higher than 150s epoch size.

3.2. Peformance comparison for ADR and RIP

The results showed that performance of $CONF_{ADR}$ and $CONF_{RIP}$ decrease by approximately 10% in terms of screening accuracy, specificity and sensitivity compared to $CONF_{PSG}$. However, the results were still satisfying for a screening approach, with accuracy, sensitivity and specificity higher than 85%. The comparison between ADR and RIP showed a performance in the same range of results whether in terms of bias, F1-Score or screening performance, which suggested the interchangeability between the two measurement techniques in a polysomnography context.

3.3. Features engineering analysis

Fig 8 presents the features weight after NCA using $CONF_{PSG}$ for an epoch size of 300s. Features from $AIRFLOW_{PSG}$ and SpO_2 carried the highest weight for classification compared to features extracted from ECG. More specifically, the features with the highest weight (> 50% of the maximum weight) were *NPRate (2):* Non-Periodic windows Rate, *LempZC (4)* : Lempel-Ziv complexity, *ODI25 (5):* number of occurrences that SpO_2 level declines at least 2% below the baseline and lasts at least 5 seconds and *VarDispMetric (3):* Variance of the Dispersion Metric calculated in every window during the epoch. *Kurtosis* (1) is a measure of the flattening of the

Table 6. Classification and AHI estimation results

Configuration	Epoch Size (s)	F1-Score (%)	Bias	Acc (%)	Sen (%)	Spe (%)
$CONF_{PSG}$	60	64 ± 17	-7.15 ± 8.41	82	100	62
	150	72 ± 16	-3.10 ± 8.88	93	100	85
	300	**74 ± 17**	**-1.27 ± 10.08**	**100**	**100**	**100**
$CONF_{ADR}$	60	59 ± 18	-5.04 ± 10.49	79	87	69
	150	67 ± 19	-0.99 ± 9.46	86	87	85
	300	**69 ± 18**	**3.47 ± 9.77**	**89**	**80**	**100**
$CONF_{RIP}$	60	61 ± 18	- 7.20 ± 9.00	82	100	62
	150	69 ± 18	-0.85 ± 7.75	89	93	85
	300	**69 ± 19**	**1.64 ± 11.55**	**89**	**87**	**92**

statistical distribution of the signal. The higher the Kurtosis, the higher the probability of abnormal presence.

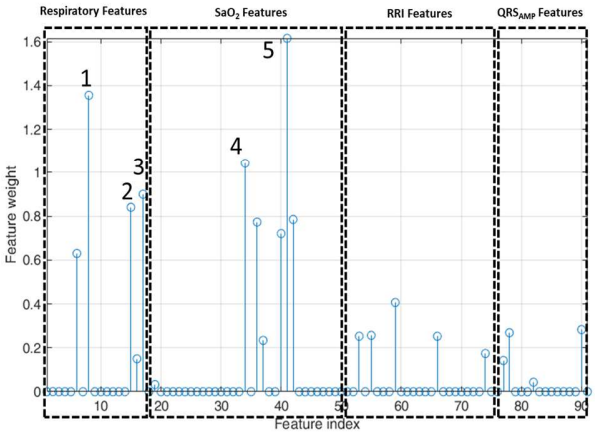

Figure 8: Features weight analysis using NCA for $CONF_{PSG}$ (EpochSize = 300s). 1: *Kurtosis*, 2: *NPRate*, 3: *VarDispMetric*, 4: *LempZC*, 5: *ODI25,* were the features with the highest weight.

Furthermore, an AHI estimation on 300s epoch using $CONF_{PSG}$ without the three ventilatory specific features (*NPRate*, *ERate* and *VarDispMetric*) had been implemented to evaluate the value of these features in the classification process. The results of this evaluation are presented in Fig 9.

A decrease of 11% in the AHI estimation accuracy was observed without these features. Specificity also dropped from 100% to 85% such as sensitivity that dropped from 100% to 93%.

These features could therefore be meaningful and relevant for improving model performance and distinguishing positive from negative epochs.

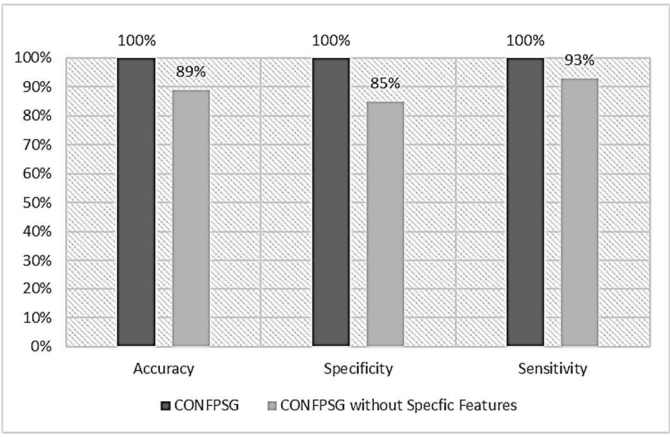

Figure 9: Results comparison between $CONF_{PSG}$ using all features and $CONF_{PSG}$ without the ventilatory specific features (*NPRate*, *ERate* and *VarDispMetric*) for a 300s epoch size.

3.4. Performance comparison with previous works

The performance comparison is presented in Table 7. The best performance results for $CONG_{PSG}$ and $CONF_{ADR}$ were compared to previous works in the literature. Several studies or devices that proposed an automatic approach for SAS screening approach were listed. This non-exhaustive selection focused on the diversity of the database, the inputs and the classification methods.

The results obtained in [9] and [33] showed great screening performance using classifier on ECG signals. However, their works used the Apnea-ECG Database, although of high quality, where two folds out three do not have apnea segments. The work presented in [32] used multimodality by using SpO_2 and ECG features and introduced the notion of model explanation. The limitations of their work are the absence of subjects with an AHI between 5 and 10 and the absence of respiratory

Table 7. Performance comparison with previous work (- stands for absence of data / information)

Study	Database	Effective	Input signals	Screening Threshold	Classifier	Acc (%)	Sen (%)	Spe (%)
[9]	Apnea-ECG	70	ECG	AHI > 5	SVM - HMM	97.1	95.8	100
[33]	Apnea-ECG	70	ECG	AHI > 5	NN - HMM	100	100	100
[32]	Own	70	SpO_2 ECG	AHI > 10	LDA	100	100	100
[34]	Own	32	PVDF	AHI > 5	-	96.2	100	91.7
[35]	Own	94	SpO_2	AHI > 15	-	81	63	96
[36]	Own	154	SpO_2	AHI > 15	-	-	88	90
[15]	Own	75	PAT	AHI > 15	-	-	92	77
[17]	Own	53	ECG 1x Acc	AHI >15	SVM	89.4	78.6	93.9
Proposed	**Own**	**28**	**$CONF_{PSG}$ $CONF_{ADR}$**	**AHI > 15**	**RUS Boosted Tree**	**100 89**	**100 80**	**100 100**

PVDF: Polyvinylidene Film-Based Sensor; PAT: Peripheral Arterial Tone; HMM: Hidden Markov Model; NN: Neural Network; Acc: Accelerometer

features, which could be problematic in case of patients with cardiac disorders where specific HRV features could then be irrelevant. The studies presented in [34], [35], [36] and [15] evaluated screening devices using different approaches such as SpO_2 or PAT (Peripheral Arterial Tone). The approach used for apnea segment detection was not necessarily based on machine learning but it was interesting to observe that the results presented in the present study were in the same range than these devices. Finally, our ADR solution using specific physiological features and a machine learning approach showed similar performance results for AHI estimation than the HealthPatch™ [17].

4. Discussion

Overnight physiological monitoring with an adaptive chest Accelerometry-Derived Respiratory index (ADR) technology using a machine learning approach provided an accurate estimate of the Apnea-Hyponea Index (AHI).

Such a screening tool is clinically relevant in comparison to the reference PSG configuration (PSG signals used in the model) and performs in the same range of accuracy than a configuration using an airflow estimation using RIP.

It is important to notice that overnight PSG data were collected in a supervised environment and sleep nurses often had to put RIP bands or nasal cannula back in place when they observed noisy signals or loss of signal. During the whole data collection, neither the thoracic nor the abdominal accelerometer had to be put back in place, which is a real advantage in terms of usage.

To our knowledge, it is the first time that a system using both a thoracic accelerometer and an abdominal accelerometer and proposing an airflow estimation is used for AHI estimation and SAS detection. It shows that an adaptive ADR system can be used at-home for SAS screening. Also, the use of alternative signals such as ADR could help reduce the failure rate of at home PSG when nasal cannula or RIP signals are unusable. Furthermore, this approach of using classical PSG signals in a machine learning process continues to promote what could be a very promising solution of semi-automatic PSG analysis and could reduce consequently the scoring time for sleep expert by detecting positive epochs with apnea events.

This study also showed that adding features built on an explanatory physiological approach has a real interest in improving classification performance. This could facilitate the physician adhesion to an automatic approach using machine learning.

At this level of validation, our model allowed to measure the AHI regardless of the type and origin of the event(s) within the 300-second segment. However, our design choices based on the use of a double thoracic and abdominal accelerometry could allow us to distinguish between the different respiratory event type. This will be one focus for our future research.

A limitation of our study is therefore the small dimension of our database. Even if it was rather well distributed in term of AHI, another clinical study has to be carried out on more subjects, in order to evaluate the reproducibility of the measure and to increase the statistical dimension.

Acknowledgments: The authors thank Pr. Renaud Tamisier, Corinne Loiodice and the team of the sleep laboratory of Grenoble Hospital for their work during the sleep study. The authors would also like to thank the team of Holi for funding and supporting this study, especially Grégoire Gerard, Damien Colas and Marc Holtzwarth.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] Lévy, P., Kohler, M., McNicholas, W. et al. Obstructive sleep apnoea syndrome. Nat Rev Dis Primers 1, 15015 (2015).

[2] Benjafield A, Ayas A, Eastwood P, Heinzer R, Ip M, Morrell M, Nunez C, Patel S, Penzel T, Pépin JL et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. The Lancet Respiratory Medicine. Volume 7, Issue 8. 2019, Pages 687-698. ISSN 2213-2600.

[3] Mannarino MR et al., "Obstructive sleep apnea syndrome," Eur. J. Intern. Med. 2012; vol. 23, no. 7, pp. 586 593.

[4] De Groote A et al. "Detection of obstructive apnea events in sleeping infants from thoracoabdominal movements". Journal of Sleep Research, 2002; 11:161-168

[5] Billiard M, Dauvilliers Y. « Les troubles du sommeil ». Elsevier Masson, 2011. 2ème édition. ISBN : 978-2-294-71025-4

[6] K. E. Bloch, "Polysomnography: A systematic review," Technol. Health. Care, 1997; vol. 5, no. 4, pp. 285–305

[7] Punjabi NM et al. "The Epidemiology of Adult Obstructive Sleep Apnea", Proc Am Thorac Soc, 2008; Vol 5. pp 136–143.

[8] Varon C et al. “A Novel Algorithm for the Automatic Detection of Sleep Apnea from Single-Lead ECG”, IEEE Transactions on Biomedical Engineering, 2015; vol. 62, no. 9.

[9] Song C et al. “ An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals”, IEEE Transactions on Biomedical Engineering, 2016; vol. 63 , no. 7.

[10] Li K et al., “A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal”, Neurocomputing, 2018; vol. 294, pp 94.

[11] Mostafa SS et al. “SpO_2 based sleep apnea detection using deep learning”, IEEE 21st International Conference on Intelligent Engineering Systems (INES), 2017.

[12] Nikkonen, S et al. “Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea”. Scientific Report, 2019; vol 9, 13200;

[13] Almazaydeh L et al. “A Neural Network System for Detection of Obstructive Sleep Apnea through SpO_2 Signal”, International Journal of Advanced Computer Science and Applications, 2012; vol. 3, no. 5.

[14] Penzel T et al. “The apneaECG database,” in Proc. Computers in Cardiology, 2000; pp. 255–258.

[15] Garg N et al. “Home-based Diagnosis of Obstructive Sleep Apnea in an Urban Population”, Journal of Clinical Sleep Medicine, 2014; 10(8):879-85.

[16] C. A. Nigro et al. “Comparison of the automatic analysis versus the manual scoring from ApneaLink device for the diagnosis of obstructive sleep apnoea syndrome,” Sleep and Breathing, 2011; vol. 15, no. 4, pp. 679–686

[17] Selvaraj N, Narasimhan R. “Automated prediction of the Apnea-Hypopnea Index using a wireless patch sensor”. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014; 1897-1900.

[18] Sànchez Morillo D et al. “An accelerometer-based device for sleep apnea screening”. IEEE Transactions on Information Technology in Biomedicine, 2010; 14-2.

[19] Bucklin CL, Das M, Luo SL. “An inexpensive accelerometer-based sleep-apnea screening technique”. IEEE Transactions, 2010; 396-399

[20] Sweeney KT et al. “Identification of sleep apnea events using Discrete Wavelet Transform of respiration, ECG and accelerometer signals”. IEEE Transactions, 2013; 10.1109

[21] Bricout A et al. “Adaptive Accelerometry Derived Respiration: Comparison with Respiratory Inductance Plethysmography during Sleep“, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019; 10.1109.

[22] Task Force of the European Society of Cardiologyandthe North American Society of Pacing and Electrophysiology. “Heart rate variability. Standards of measurement, physiological interpretation, and clinical use”. Eur. Heart J. 17, 1996; (3), 354–381

[23] A Eberhard A et al. “Comparison between the respiratory inductance plethysmography signal derivative and the airflow signal”,Adv. Exp. Med. Biol, 2001; vol. 499, pp. 489-494.

[24] Carry, P.-Y., Baconnier, P., Eberhard, A., Cotte, P., & Benchetrit, G. (1997). Evaluation of Respiratory Inductive Plethysmography: Accuracy for Analysis of Respiratory Waveforms. Chest, 111(4), 910-915

[25] D. Singh, V et al. “Sampling frequency of the RR-interval time-series for spectral analysis of the heart rate variability”. Journal of Medical Engineering Technology, 2004; 28(6):263–272.

[26] J. Pan, W. J. Tompkins, "A real-time QRS detection algorithm", IEEE Trans. Biomed. Eng. 1985; vol. BME-32, no. 3, pp. 230-236.

[27] Sedghamiz, Hooman, “Matlab Implementation of Pan Tompkins ECG QRS detector”, 2014 10.13140/RG.2.2.14202.59841.

[28] Xie B et al. “Real-Time Sleep Apnea Detection by Classifier Combination”, IEEE Transactions on Information Technology in Biomedicine, 2012; vol. 16, no.3.

[29] Tiago H. Falk et al. “Augmentative Communication Based on Realtime Vocal Cord Vibration Detection”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2010; vol. 18, no. 2;

[30] Goldberger J et al. “Neighbourhood components analysis”, Advances in Neural Information Processing Systems 17, 2005; pp. 513–520, MIT Press.

[31] Seiffert C et al. “RUSBoost: A Hybrid Approach to Alleviating Class Imbalance”, IEEE Transactions on Systems, Man, And Cybernetics—Part A: Systems and Humans, 2010; vol. 40, no. 1.

[32] Ravelo-García AG et al. “Oxygen Saturation and RR Intervals Feature Selection for Sleep Apnea Detection”, Entropy, 2015; vol.17, pp. 2932-2957.

[33] Li K et al. “A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal”, Neurocomputing, 2018; vol 294, pp. 94–101.

[34] Hwang SH et al. “Unconstrained Sleep Apnea Monitoring Using Polyvinylidene Fluoride Film-Based Sensor”, IEEE Transactions on Biomedical Engineering, 2014; vol. 61, no. 7;

[35] Jobin V et al. “Predictive value of automated oxygen saturation analysis for the diagnosis and treatment of

obstructive sleep apnoea in a home-based setting", Thorax, 2007; 62(5):422-7.

[36] Nigro C, et al. "Validation of the WristOx 3100™ oximeter for the diagnosis of sleep apnea/hypopnea syndrome", Sleep Breath, 2009; 13(2):127-36.