Unsupervised Artefact Detection and Screening Using Emfit Sensor in Patients With Sleep Apnea

Dorien Huymans[1,2], Bertien Buyse[3], Dries Testelmans[3], Sabine Van Huffel[1,2], Carolina Varon[1,2]

[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics, Leuven, Belgium
[2] imec, Leuven, Belgium
[3] UZ Leuven, Department of Pneumology, Leuven, Belgium

Abstract

Sleep apnea is one of the most common sleep disorders. As sleep apnea is associated to adverse health outcomes, early screening is promoted through unobtrusive, cheap and simple systems for sleep monitoring. A commercial pressure sensor meeting these requirements is the Emfit QS, which was integrated in a bed of a specialized sleep center. The sensor is pressure based and highly sensitive to movement. This causes artefacts of different morphologies in the signal. An unsupervised artefact detection method was developed to avoid burdensome manual labelling of artefacts in the signal and enabling further analysis. Moreover, the percentage of detected artefacts was useful for assessment of the sleep apnea severity as movements partially originate from apneic arousals. Severe sleep apnea patients could be identified with a sensitivity of 80% and a specificity of 87%.

The proposed approach offers an ambivalent tool for artefact detection and unobtrusive screening of sleep apnea patients at home.

1. Introduction

Sleep apnea is one of the most common sleep disorders with an expected increase in prevalence [1]. It is characterized by breathing cessations causing frequent arousals from sleep. It heavily disturbs the patient's night sleep, leading to a range of health problems such as daytime sleepiness and cardiovascular diseases.

The gold standard method for diagnosis is a full-night polysomnography (PSG) in a specialized sleep centre [2]. However, the extensive sensor set-up causes patient discomfort and does not replicate a normal sleeping environment. Moreover, the PSG procedure requires special training for analysis, is costly and burdensome. Sleep centres often have a limited capacity as well. Therefore, simple, cheap and unobtrusive measurement systems are desired. These can be used in a home-environment to complement the current clinical practice such as pre-clinical screening. With these characteristics, the systems would allow long-term monitoring as well.

The Emfit is a pressure sensor meeting these requirements. From this pressure signal, a respiratory signal and ballistocardiography (BCG) can be derived. BCG is an unobtrusive measurement of the body's recoil caused by cardiovascular pulsation. Tenhunen et al. [3] defined several Emfit-derived parameters and found correlations with AHI. A prerequisite for their method was a visual scoring of breathing patterns into nine categories. The authors did not develop an automated detection of breathing patterns and manual annotation was still required. Moreover, the Emfit is highly sensitive to movement. This causes artefacts of different morphologies in the signal. Bruser et al. developed an algorithm [4] to separate Emfit BCG epochs into epochs with normal sinus rhythm, atrial fibrillation and artefacts using common supervised machine learning algorithms.

Here, an unsupervised artefact detection method was developed to avoid burdensome manual labelling of artefacts in the signal and enabling further analysis. Furthermore, the current study explored the use of Emfit as an unobtrusive screening tool of patients at risk of sleep apnea based on the percentage of detected artefacts.

2. Methods

2.1. Data

The Emfit QS is a commercially available pressure sensor. The sensor (542mm $\times$ 70mm $\times$ 1.4mm) was placed right beneath the mattress cover under the patient's chest area to minimize the distance to the heart and maximize signal quality. The data was sampled at 100Hz.

The Emfit sensor and PSG recorded simultaneously data of 37 patients (33 male, age: 48.4 ± 11.4 years, BMI:

$30.4 \pm 7.5 \frac{kg}{m^2}$, AHI: 28.0 ± 24.0) during sleep diagnosis in UZ Leuven. Overnight PSG signals were annotated by sleep specialists according to the AASM 2012 scoring rules [2] to derive the AHI. Three patients did not suffer from sleep apnea, 9 suffered from mild apnea, 13 from moderate apnea and 12 from severe apnea.

2.2. Emfit Preprocessing

After subtraction of the mean value, the raw pressure signal of the Emfit sensor was filtered with a Butterworth bandpass filter to obtain both the respiration (0.08-1Hz) and the BCG (6-16Hz) signal. The respiratory signal was resampled at 8Hz and the BCG signal at 50Hz.

Next, discrete wavelet transforms were applied on all three signals, namely the raw pressure signal, the respiration and the BCG, to capture time-frequency domain information. A Daubechies 1 (*db1* or Haar) wavelet was applied on the pressure signal to accentuate steep changes in the signal. The signal was decomposed until level 8, i.e. [0.2, 0.4]Hz. The respiratory signal was approximated with a *db4* wavelet (until level 4, [0.25, 0.5]Hz) and the BCG with *db6* (until level 2, [6.25, 12.5]Hz). The respective wavelet shapes were chosen for its resemblance with the natural wave shape. A total of 17 signals (original signals and decompositions) were used for the subsequent feature extraction step.

2.3. Feature Extraction

Time domain features were derived from both the untransformed signals and the wavelet decomposed signals (see Table 1). A feature window of 10s was applied to accurately locate artefacts and include two to three breaths from the respiration signal. Features 1-5 are expected to be affected by outliers caused by artefacts. To capture the regularity of the signal, the Kurtosis of the Autocorrelation function and Shannon Entropy were derived. Furthermore, the peak-to-peak (PP) amplitude (PP $= \mathbf{max}(\mathbf{x}) - \mathbf{min}(\mathbf{x})$) of the segment should be very large in the presence of a movement artefact. The next features are based on a PP series within the feature window. The window is split into 3 equal subsegments over which PP is calculated, resulting in $\mathbf{PP_3}$ [4].

Features were standardized per subject to zero mean and unit variance in order to map the different features within similar ranges and to account for inter-subject variability. Highly correlated features from Table 1 were removed using Pearson's linear correlation coefficient and a threshold of 90%. Feature values were transformed by Euclidean norm normalization, to decrease the effect of extreme values.

Table 1. Features extracted from 10s windows.

	Feature
1-3	Mean, Variance (=Var), Standard Deviation (=Std),
4-5	Kurtosis, Skewness
6-7	Kurtosis of Autocorrelation, Shannon Entropy
8	Peak-to-Peak Amplitude
9	Maximum ($\mathbf{PP_3}$) / mean ($\mathbf{PP_3}$)
10-11	Var ($\mathbf{PP_3}$) (=peakVar), Std ($\mathbf{PP_3}$)
12	[10%, 25%, 50%, 75% 90%] ($\mathbf{PP_3}$)
13	Inter Quartile Range ($\mathbf{PP_3}$)
14	Inter Decile Range ($\mathbf{PP_3}$)
15	Median Absolute Deviation ($\mathbf{PP_3}$)

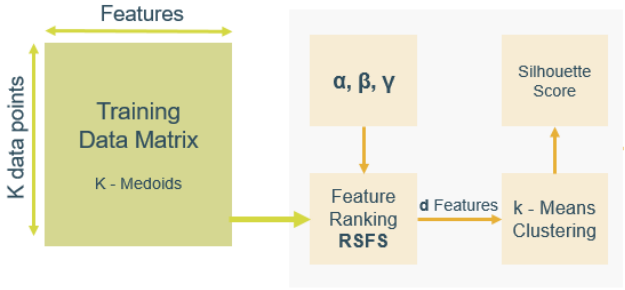

Figure 1. Algorithm for unsupervised feature selection.

2.4. Unsupervised Feature Selection

An algorithmic pipeline for unsupervised feature selection was developed as depicted in Fig. 1. The input was a K-medoids clustering to reduce the dataset and select training points [5], using a number of clusters $K_{med} = 2000$ and the Mahalanobis distance metric. This selection served as the input for an unsupervised feature selection framework based on Robust Spectral learning (RSFS) [6]. Parameters α, β and γ of the RSFS were optimised with grid search over equispaced values in logarithmic scale from -3 to 3 for each parameter. For every set of parameters a feature ranking was calculated with RSFS. After ranking of all input features with a set of α, β and γ, a number d of top ranked features was selected. Subsequently, a k-means clustering in a d dimensional space was performed 20 times using squared euclidean distance and random initialisation. The performance of every k-means clustering step was evaluated by the overall average silhouette score. The entire pipeline from parameter selection to clustering has been repeated for $d = [3, 5, 7]$ features and $k = 2$ clusters. The output of optimization were α, β and γ for feature ranking and the top ranked features (type and amount d).

2.5. Clustering of Artefacts and Arousals

The optimised parameters of the preceding step were applied to cluster the training points in two clusters. The centroids of these trained clusters acted as target points for the testing data. Every test data point was mapped to the closest centroid in Euclidean distance to determine its associated cluster. The characteristics of the clusters were analysed based on their feature values and running a pairwise Mann-Whitney U test.

It was assumed that one cluster would contain *clean* and the other *distorted* data segments due to the nature of the feature set. During sleep it was expected that the smallest cluster of mapped test data contained artefacted segments, mainly originating from movement or apneic arousals (further referred to as *artefact* cluster). It was hypothesised that the percentage of data segments belonging to the *artefact* cluster increases with AHI as more movement and arousals would be detected. Patients were grouped according to standard sleep apnea classes based on AHI. Distributions of *artefact* cluster sizes were derived for respective classes. A Krukal-Wallis test with Bonferroni correction was applied to test which classes stochastically dominate the others. To discriminate severe sleep apnea (AHI$\geq$ 30) patients from other sleep apnea patients (AHI$<$ 30), a receiver operating characteristic (ROC) curve analysis was performed based on *artefact* cluster size.

2.6. Screening

In order to apply the previously described method for screening, a leave-one-patient-out (LOO) approach was taken. A K-medoids training sample selection was made based on 36 patients. The optimised features were extracted from the training sample selection followed by k-means clustering. The two centroids of the resulting clusters acted as reference points for *artefact* and *clean* data of the left out patient. The percentual *artefact* cluster size of every LOO patient was assigned to the sleep apnea severity class according to the patients' AHI.

3. Results

3.1. Clustering

The unsupervised feature selection was based on an algorithmic pipeline with optimization of different parameters. As optimal parameter sets slightly varied, feature ranking varied as well. However, certain subsets of features were favoured in the top 10 ranked features over several K-medoids iterations. Within 24 iterations, 28 unique features occurred in the top 10. Features *resp db4 level1 peakVar*, *pressure db1 level4 peakVar* and *pressure db1 level3 peakVar* were chosen as a performant subset of features and occurred respectively 37.5%, 75.0% and 79.2% in top 10 features.

After ranking all input features with a set of α, β and γ, a number d of top ranked features was selected. Subsequently, a k-means clustering in a d dimensional space was performed. A low number of features was preferred. The entire pipeline from parameter selection to clustering was repeated as well for $k = [3, 4, 5, 6]$. With a higher number of clusters $k > 2$, the overall average silhouette score decreased, as probably the natural existing clusters are broken up into multiple ones.

Based on 24 iterations of the pipeline for training samples selection and subsequent parameter selection, the following parameters were found to consistently provide a good performance: $\alpha = 100$, $\beta = 1000$, $\gamma = 0.001$; $d = 3$; *resp db4 level1 peakVar*, *pressure db1 level4 peakVar* and *pressure db1 level3 peakVar* (see Table 1); $k = 2$.

The overall average silhouette score of both clusters of training samples was 0.83. Cluster 1 contained the majority of the training samples with a highly varying silhouette score. In contrary, cluster 2 was a very well defined cluster and should contain samples with similar characteristics. The different cluster sizes can be explained by the K-medoids selection of training samples which captured different characteristics of the underlying data. Training samples related to different types of distortions exhibited a varying morphology. Every morphology needed to be represented by a different K-medoids centroid. Therefore, the majority of K-medoid centroids were related to distortions.

Mapping the test data to the trained centroids resulted in an overall average silhouette score of both clusters of 0.92. The increase in score was due to a larger bulk of clean segments within the test data. One cluster containing higher values of features highlighting peak variations was labelled as *artefact* cluster. The other cluster characterized stable segments without intermittent peaks and was labelled as *clean* cluster. The difference in data distribution between clusters was high (Mann-Whitney U test, $p < 0.001$), indicating that parameters were well optimised to make a distinction between *artefact* and *clean* data.

3.2. Screening

Several outliers were present in the distributions of *mild*, *moderate* and *severe* classes. Three out of four outliers belonged to patients with a BMI exceeding the 90th percentile of BMI ($= 42.02 \frac{kg}{m^2}$). A very high BMI tends to saturate the pressure signal and decrease signal quality. Therefore, patients with a BMI exceeding the the 90th percentile were removed from the analysis (Fig. 2). With a Kruskal-Wallis test, significant differences were observed between *mild* - *severe* class ($p < 0.05$) and *moderate* - *severe* class ($p < 0.05$).

The optimal threshold to identify patients at high risk

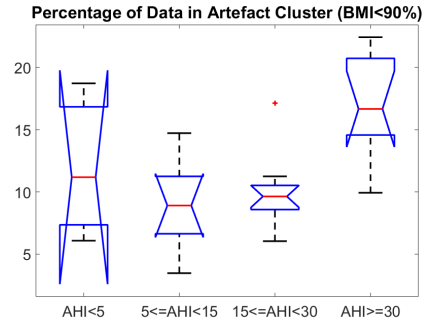

Figure 2. Percentage of data assigned to *artefact* cluster per patient, subdivided per apnea severity class.

of suffering from sleep apnea was determined by ROC analysis, with an area under curve of 0.90. A threshold of 14.57% artefacts resulted in a sensitivity of 80% and specificity of 87%. False positive results were related to a decreased signal quality, a %artefacts close to the decision threshold or a patient nearly at risk of severe sleep apnea. False negative results could be assigned to patients with a very high number of obstructive hypopneas (Hobs). In case Hobs were not followed by arousals, these are difficult to detect by the type of features used in the algorithm.

4. Discussion

An unsupervised artefact detection method for a pressure sensor was developed. Additionally, the approach presented here offers a tool for unobtrusive screening of patients at risk of severe sleep apnea at home with a relatively cheap sensor. These patients should be referred for further research in a sleep clinic and can be prioritized on the waiting lists.

A limitation of the tool is the exclusion of patients with a large BMI as it causes the signal to saturate and decrease the signal quality. In this study a threshold of $42.02 \frac{kg}{m^2}$ has been set, meaning an exclusion of 10.8% of all patients of which 50% suffered from severe apnea.

Another limitation is the difficulty of detecting apneic events without arousals, often seen with obstructive hypoapneas (Hobs) or with periodic breathing. The change in signal amplitude is limited or slowly waxing and waning. The window length of 10s is too small to detect this and the type of features are rather tuned for steep changes.

Artefacts captured both movement due to position changes as well as limb movements and arousals. Also, the %artefacts was calculated over recording time, including sleep and wake stages instead of total sleeping time. Therefore, %artefacts could not be correlated to the AHI.

Further research is aimed at screening patients suffering from mild and moderate sleep apnea using more refined features based on respiration rate and heart rate. Additionally, spiking patterns in the BCG band were reported temporally related to upper airway obstruction, which could be further investigated for apnea detection [7].

5. Conclusion

The Emfit pressure sensor was explored in its potential for sleep apnea screening. An unsupervised algorithmic pipeline based on clustering was developed to detect artefacts and relate these artefacts to sleep apnea classes. The percentage of artefacts in the data was a useful parameter to target severe sleep apnea patients.

Acknowledgements

Agentschap Innoveren en Ondernemen (VLAIO): 150466: OSA+ ; Agentschap voor Innovatie door Wetenschap en Technologie (IWT): O&O HBC 2016 0184 eWatch ; imec funds 2017 ; European Research Council: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.; Carolina Varon is a postdoctoral fellow of the Research Foundation-Flanders (FWO).

References

[1] Peppard PE, et al. Increased prevalence of sleep-disordered breathing in adults. Am J Epidemiol 2013;177 9:1006–14.

[2] Berry R, et al. Rules for scoring respiratory events in sleep: Update of the 2007 aasm manual for the scoring of sleep and associated events. JCSM 11 2012;8(5):597–619.

[3] Tenhunen M, et al. Emfit movement sensor in evaluating nocturnal breathing. Respir Physiol Neurobiol 2013; 187(2):183–9.

[4] Bruser C, et al. Automatic detection of atrial fibrillation in cardiac vibration signals. IEEE JBHI 2013;17(1):162–171.

[5] Varon C, et al. Noise level estimation for model selection in kernel pca denoising. IEEE Trans Neural Netw Learn Syst 2015;26(11):2650–2663.

[6] Shi L, et al. Robust spectral learning for unsupervised feature selection. In 2014 IEEE ICDM. 2014; 977–982.

[7] Kirjavainen T, et al. Respiratory challenge induces high frequency spiking on the static charge sensitive bed (scsb). Eur Respir J 1996;9(9):1810–1815.

Address for correspondence:

Dorien Huysmans
ESAT/STADIUS/KU Leuven
Kasteelpark Arenberg 10, bus 2446, 3001 Leuven, Belgium
dorien.huysmans@esat.kuleuven.be