Novel method for detection of Sleep Apnoea using respiration signals

Kristine Carmes[1], Lykke Kempfner[2], Helge Bjarup Dissing Sorensen[1] and Poul Jennum[2,3]

Abstract— **Polysomnography (PSG) studies are considered the "gold standard" for the diagnosis of Sleep Apnoea (SA). Identifying cessations of breathing from long-lasting PSG recordings manually is a labour-intensive and time-consuming task for sleep specialist, associated with inter-scorer variability. In this study a simplified, semi-automatic, three-channel method for detection of SA patients is proposed in order to increase analysis reliability and diagnostic accuracy in the clinic. The method is based on characteristic features, such as respiration stoppages pr. hour and the total number of oxygen desaturations > 3%, extracted from the thorax and abdomen respiration effort belts, and the oxyhemoglobin saturation (SaO_2), fed to an Elastic Net classifier and validated according to American Academy of Sleep Medicine (AASM) using the patients' AHI value. The method was applied to 109 patient recordings and resulted in a very high SA classification with accuracy of 97.9%. The proposed method reduce the time spent on manual analysis of respiration stoppages and the inter- and intra-scorer variability, and may serve as an alternative screening method for SA.**

I. Introduction

SA (obstructive and central) is a common sleep disorder. Obstructive sleep apnoea (OSA) is characterized by complete or partial repetitive upper airway obstruction and is the most common type of SA: it affects more than 2% of children, more than 5% of adults and is associated with significant morbidity, mortality and societal burden. [1]

The severity of SA is defined by the number of apnoea and hypopnea events, lasting longer than 10 seconds, per hour of sleep (apnoea/hypopnea index–AHI). [2] SA is usually diagnosed when the patient has AHI $\geq$ 5 events/h accompanied with excessive daytime sleepiness and/or cardiovascular morbidity.

PSG studies, a routine examination performed on numerous sleep laboratories throughout the world, are considered the "gold standard" for the diagnosis of SA and other sleep disorders. Identifying cessations of breathing from long-lasting PSG recordings manually is a labour-intensive and time-consuming and associated with inter- and intra-scorer variability [3,4], making the diagnosis of SA underestimated considering general health check-ups. An efficient and more accessible computer-assisted system for detection of SA would thus offer a powerful clinical diagnostic tool from the perspective of analysis reliability, diagnostic accuracy, efficiency and economic burden.

Several automatic methods have been proposed for feature extraction, reduction and recognition of cessation of breathing, and assessed on various biomedical signals, with very good results [5-8]. However, most studies only included patients with moderate and severe SA, since mild SA most often do not require treatment. Also, most methods are more advanced than needed and designed for real-time monitoring of SA. We propose a simplified, efficient, semi-automatic, three-channel method for detection of SA patients, based on characteristic features, fed to an elastic net and validated according to international standards [2].

II. Materials

A. Data selection

A total of 109 subjects, divided into six different patient groups, were included in this study: 19 healthy subjects, 19 patients with increased periodic leg movements (PLM), 20 patients with idiopathic Rapid eye movement (REM) behaviour disorder (iRBD), 18 SA patients, 20 SA patients with PLM and 13 SA patients with RBD, see Table I for demographic details on the patients and Table II and III for AHI distribution.

TABLE I: **Demographic patient details (mean ± 95% CI)**

Group	Subjects (n)	Age (year)	BMI (kg/m^2)	Females (n)
Healthy	19	45 ± 15.5	24.4 ± 5.1	14
PLM	19	56.3 ± 11.7	25 ± 3.9	9
iRBD	20	61 ± 8	25.9 ± 5.3	4
SA	18	57.3 ± 12.7	30.5 ± 5.8	5
SA/PLM	20	60.8 ± 10	27 ± 4.5	7
SA/iRBD	13	63 ± 8.3	26.9 ± 3.6	2
Total	109			

TABLE II: **Subject AHI distribution**

AHI	# train subjects	# validation subjects
AHI 5-15 (mild)	15	4
AHI 15-30 (moderate)	13	2
AHI > 30 (severe)	10	7

TABLE III: **Group AHI distribution**

Group	AHI (mean±95% CI)
Healthy	0.7 ± 1
PLM	1.5 ± 1.7
iRBD	1.2 ± 1.7
SA	44.3 ± 23.9
SA/PLM	20.5 ± 10.4
SA/iRBD	16.3 ± 14.1

Patient groups such as PLM and iRBD patients are included in the study in order to design a model as generic as possible.

[1] Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark.

[2] Danish Centre for Sleep Medicine, Glostrup University Hospital, Glostrup, Denmark.

[3] Center for Healthy Ageing, University of Copenhagen, Copenhagen, Denmark.

The two groups form, together with the healthy subjects, a SA negative group. Due to the prevalence of the disorders the gender distribution in the iRBD and SA groups are mismatched. The enrolled subjects were all diagnosed and scored by sleep specialists according to the international AASM standard. That is, in this study subjects with an AHI above 5 are considered to have SA even though they do not need treatment. Also, subjects with PLM index above 5 (PLMI > 5) are considered to have PLM.

B. Polysomnographic recording

All 109 patients underwent a diagnostic full-night polysomnography (PSG), recorded at the Danish Centre for Sleep Medicine, Glostrup hospital, Denmark, involving 6 channel electroencephalography (EEG) (F_3-A_2, F_4-A_1, C_3-A_2, C_4-A_1, O_1-A_2 and O_2- A_1), electromyography (EMG) (chin, tibialis left, tibialis right), horizontal and vertical electrooculography (EOG), electrocardiography (ECG), respiration (nasal flow, thorax and abdomen respiration effort belts) and finally SaO_2. Approximately seven to eight hours of sleep per subject were available when using data from lights-off to lights-on. The data were not visually inspected for electrode discontinuities and artefacts. Specialists scored the sleep stages manually, according to the standard from the AASM [2].

In this study, the two respiration signals; thorax and abdomen respiration effort belts, the SaO_2, together with the manually scored hypnograms, were used in order to distinguish SA patients from the SA negative group. The two respiration effort signals, and the SaO_2 signal were recorded using sampling frequencies of 256Hz and 1Hz respectively, and recorded using filter settings within the AASM recommendations (0.1 Hz – 15 Hz).

III. Methodology

The detection process involves a processing step, followed by feature extraction and reduction. Finally, each recording is then classified as normal or abnormal.

A. Biomedical signal pre-processing

All data between lights-off and lights-on was pre-processed and used. The two respiration effort signals were low pass filtered using a 4th order Butterworth filter with a 3dB cut-off frequency 0.7Hz [8]. To reduce the data size the signals were down sampled to 4Hz using a low pass FIR filter. In order to avoid interference from high amplitude outliers, outliers were removed from both signals using the following histogram method. The data was binned into 50 containers according to amplitude. Samples within containers containing less than 0.3% of the total amount of data were removed. The number of bins and the removal threshold was found by trial and error. Furthermore, the data is normalized using the min-max method, scaling the data between $[-1,1]$, cf. eq. (1).

$$\bar{X} = 2 \cdot \frac{X - \min(X)}{\max(X) - \min(X)} - 1 \qquad (1)$$

The SaO_2 signal was restricted to saturation values between 70%-100%, since values below 70% are unreliable due to the recording equipment.

B. Feature extraction

Following the pre-processing steps, SA characteristic features from the respiration signals were calculated from the manually scored sleep epochs.

1) Respiration Stop Index (RSI): This feature is inspired by the AHI, and is therefore an estimate of the number of respiration stops longer than 10 seconds per hour. First, positive peaks above a pre-defined threshold (K) are detected, K = 0.61. This is executed by comparing the amplitude value of each sample with a reference window containing 5 samples. The threshold and window size are both chosen based on a trial and error method. Second, the distance between each detected positive peak larger than 10 seconds is stored as a respiration stop, and the average number of respiration stops per hour is saved as the RSI feature. In Fig.1 the peak detection is illustrated.

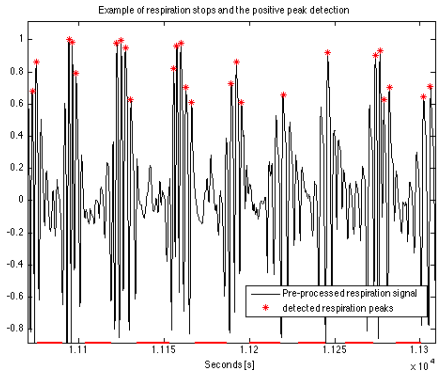

Fig. 1. An example of respiration stops in the respiration signal from the abdominal effort belt and the positive peaks detected above a threshold of K = 0.61. In this segment 8 respiration stops were detected (red lines).

2) Amount of REM and NREM sleep, and number of transitions: These three features are extracted from the manually scored hypnograms. Patients with SA have extremely fragmented and poor quality sleep, hence they do not stay in one sleep stage as long as healthy subjects due to full or partial awakenings caused by the disorder.

3) *Frequency based features:* Segments with normal, slow, rapid or no respiration are characterized by different frequencies; hence, the total energy in these different bands may distinguish healthy subjects from apnoea patients. The total energy, cf. eq. (2), in four different frequency bands, $j = 1, \ldots, 4$, were extracted from the two pre-processed respiration effort signals using 4th order band pass Butterworth filters:

- Ultra low frequency (ULF): 0.00001 – 0.013Hz
- Very low frequency (VLF): 0.013 – 0.0375Hz
- Low frequency (LF): 0.0375 – 0.06Hz
- High frequency (HF): 0.17 – 0.28Hz

The signals were segmented into N number of 30 seconds epochs after filtering. The total energy of each band is given by:

$$E_j = \sum_{n=1}^{N} |y_j|^2 \quad (2)$$

4) Number of desaturations greater than 3%: This feature is extracted from the restricted SaO_2 signal. First, each sample is investigated in relation to a reference window of 15 prior samples. All samples representing a drop in saturation compared to the median saturation in the reference window are stored. Second, the degree of desaturation is calculated and the total number of desaturations greater than 3% is saved as a feature.

C. Feature reduction and classification

The dataset was divided into training, test and validation sets. One fourth of the 109 patients, total of 28 subjects, were randomly selected within each patient group and saved for later validation. The remaining 81 subjects were used for feature investigation and selection, determination of parameters, threshold values and more.

For feature reduction and classification a penalized least squares method, called the elastic net regularized regression [9], was applied. Unlike traditional classifiers, such as Support Vector Machine (SVM) and Gaussian Mixture Model (GMM), the elastic net performs simultaneous automatic variable selection, continuous shrinkage and grouping of correlated variables [9]. The elastic net solution is a convex combination of both the lasso and ridge penalty. For an α strictly between 0 and 1, elastic net solves the optimization problem:

$$\min_{\beta_0,\beta} \left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right) \quad (3)$$

where

$$P_\alpha(\beta) = \frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 = \sum_{j=1}^{p} \left(\frac{1-\alpha}{2} \beta_j^2 + \alpha |\beta_j| \right). \quad (4)$$

N is the number of observations, y_i is the response at observation i, x_i is the data at observation i, λ is a positive regularization parameter and the parameters β and β_0 are scalar. The function $P_\alpha(\beta)$ is the elastic net penalty term, which is a convex combination of the lasso and ridge penalty. When α = 1, the elastic net becomes simple ridge regression. [9] The 10-fold-cross validation was applied to the training data to obtain optimum alpha value in terms of performance. The final model was re-trained using the complete training set and optimum alpha, and further assessed on the independent validation set. The model output is the class probability. For validation purposes the class threshold for SA was set at $\geq 50\%$ probability of SA. The final output, *apnoea patient or not*, is validated according to the patients' AHI value and not the manually scored hypnograms or temporal placement of the respiration stoppages.

The area under the receiver operating characteristic (ROC) curve, called AUC, the sensitivity, cf. eq. (5) and specificity, cf. eq. (6) was computed and used as performance measures.

$$SE = 100 \cdot \frac{TP}{TP+FN} \quad (5)$$

$$SP = 100 \cdot \frac{TN}{FP+TN} \quad (6)$$

IV. Results

The generalization performance of a classifier depends primarily, on the selection of good features, i.e., features resulting in maximum separation between the classes. The feature reduction was performed, using the elastic net algorithm during the training of the classification model. The optimal parameter α was found at 0.90 and a total of 13 features were found to be of importance to the classification, cf. Table IV.

TABLE IV: **Unsorted list of the most optimal feature set**

#	Signal channel	Selected feature
1	Abdominal effort belt	RSI
2-5	Abdominal effort belt	ULF, VLF, LF, HF
6	Hypnogram	Amount of REM sleep
7	Hypnogram	Amount of NREM sleep
8	Hypnogram	Number of transitions
9	SaO_2	Number of desaturations > 3%
10	Thoracic effort belt	RSI
11-13	Thoracic effort belt	ULF, VLF, HF

The most optimal feature was found to be feature 9, the number of desaturations > 3%, cf. Fig. 2. This feature, by itself, shows a performance of 97.7% specificity and a sensitivity depending on the severity of the SA: 5 ≤ AHI < 15 – 46.7%, 15 ≤ AHI < 30 – 75% and AHI ≥ 30 – 100%.

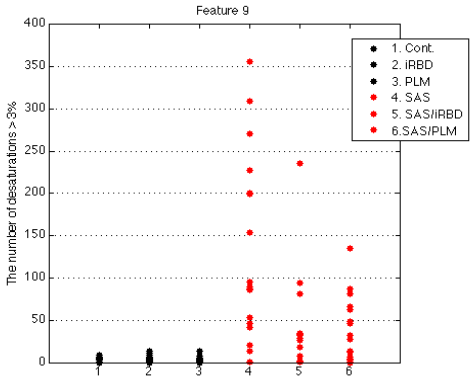

Fig. 2. Illustration of the discriminative property of feature 9 for all 109 subjects, cf. Table IV. 1: control. 2: iRBD. 3: PLM. 4: SA. 5: SA/iRBD. 6: SA/PLM. BLACK: SA negative group. RED: SA positive group.

In Fig. 3 the correlation between the AHI determined by a specialist via manually inspection and feature 9 is illustrated, the correlation value is $\rho = 0.87$.

The model performance obtained with the optimal feature subset shows that patients with SA could be separated from subjects without, using the presented method, with a very high accuracy, AUC of 97.9%, sensitivity depending on the severity of SA and a very high specificity on the test data, cf.

Table V. The unseen validation data, consisting of 5 controls, 5 PLMs, 5 iRBDs, 5 SA, 5 SA/PLM and 3 SA/iRBD patients, was classified with a high accuracy, AUC of 100%, a sensitivity of 100% and a specificity of 100%.

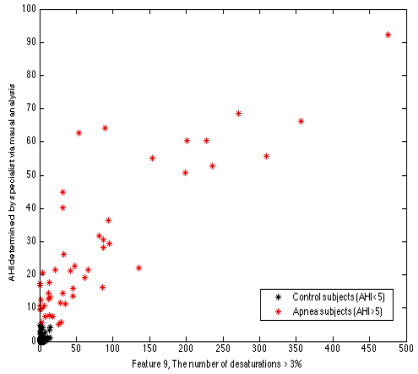

Fig. 3. The correlation between the AHI determined by a specialist via manually inspection and feature 9. BLACK: SA negative group. RED: SA positive group.

TABLE V: **Classification performance (correctly classified/#subjects)**

	Test data		Unseen validation data	
AUC	**97.9%**		**100%**	
Sensitivity	86.8%	(33 of 38)	100%	(13 of 13)
Sensitivity AHI 5-15	**66.7%**	**(10 of 15)**	**100%**	**(4 of 4)**
Sensitivity AHI 15-30	**100%**	**(13 of 13)**	**100%**	**(2 of 2)**
Sensitivity AHI > 30	**100%**	**(10 of 10)**	**100%**	**(7 of 7)**
Specificity	95.4%	(41 of 43)	100%	(15 of 15)

V. Discussion

We found that our method can separate patients with SA from subjects without, with an accuracy of 97.9% on the test data and 100% on the unseen validation data, and that feature 9 correlates well with the AHI, $\rho = 0.87$. These results are comparable or better than former methods based on respiration signals and electrophysiological signals [5-8].

The present method is robust and the model highly accurate for patients with moderate and severe SA, AHI > 15. All misclassified SA patients has an AHI lower than 12.6. The relatively low AHI makes it difficult to distinguish these subjects from the SA negative group. The test results show that the patient group SA/iRBD is the most difficult to classify. 4 of 13 patients in this group were misclassified. An explanation of the difference in performance between the different groups could be the difficulties associated with manually scoring of sleep. Due to the disorder, iRBD, it may be difficult to distinguish between REM sleep and awake, consequently the misclassified REM sleep is not investigated for respiration stoppages. The proposed method reduce the time spent on manual analysis of respiration stoppages, from approximate 1 hour to 1 minute, using a standard 2.4GHz MacBook pro and MATLAB R2011b, and reduce the inter- and intra-scorer variability due to the automation.

Limitations of the study include the use of data recorded at only one sleep lab with only one type of sensor. The method needs to be tested on external data recorded with different sensors and amplifiers in order to validate its general performance and usability. Furthermore the method is dependent on the manually scored hypnograms, which limits the use as a potential screening method for SA.

VI. Conclusion

The semi-automatic method is able to achieve excellent and comparable performance using respiration signals, SaO_2 and the manually scored hypnograms, AUC of 97.9%, which is easy to measure and clinically acceptable. The proposed method reduce the time spent on manual analysis of respiration stoppages, from approximate 1 hour to 1 minute, and the inter- and intra-scorer variability, and may, together with a future automatic sleep stager, serve as an alternative screening method for SA.

VII. Acknowledgement

Laura Graves Ponsaing and Rune Frandsen from the Danish Centre for Sleep Medicine, Glostrup Hospital, Denmark, has been a great support to this work regarding the choice of SA and RBD patients.

References

[1] P. Jennum, J. Kjellberg. *Health, social and economical consequences of sleep-disordered breathing: a controlled national study*. Thorax. 2011, Jul;66(7):560-6.

[2] C. Iber, S. Ancoli-Israel, A. Chesson, and S. Quan, *The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification*, 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.

[3] U. J. Magalang, N. H. Chen, P. A. Cistulli PA, et al., *Agreement in the scoring of respiratory events and sleep among international sleep centers.* Sleep, 2012, Volume 35, Issue Suppl. S, pp. A425.

[4] S. T. Kuna, R. Benca, C. A. Kushida, et al., *Agreement in computer-assisted manual scoring of polysomnograms across sleep centers*. Sleep. 2013, Volume 36, Issue 4, pp. 583-589.

[5] E. Goldshtein, A. Tarasiuk, and Y. Zigel, *Automatic Detection of Obstructive Sleep Apnea Using Speech Signals,* IEEE Transactions on biomedical engineering 2011, Volume 58, Issue 5, pp. 1373-1382.

[6] L. Almazaydeh, K. Elleithy, M. Faezipour, *Detection of Obstructive Sleep Apnea Through ECG Signal Features,* EIT, 2012 IEEE International Conference on 6-8 May 2012, pp.1 – 6.

[7] Z. Moussavi and A. Yadollahi, *Acoustic obstructive sleep apnea detection,* Biology Society: Engineering the Future of Biomedicine, EMBC 2009 — 2009, Volume 2009, pp. 7110-7113.

[8] N. Selvaraj and R. Narasimhan, *Detection of Sleep Apnea on a Per-Second Basis using Respiratory Signals*, 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3-7 July, 2013.

[9] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net,* J. R. Statist. Soc. B (2005). 67, Part2, pp. 301-320.