34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Apnea-hypopnea index estimation from spectral analysis of airflow recordings

Gonzalo C. Gutiérrez-Tobal, Roberto Hornero, *Senior Member, IEEE*, Daniel Álvarez, *Member, IEEE*, J. Victor Marcos, Carlos Gómez*, *Member*, *IEEE* and Félix del Campo

Abstract— **This study focuses on the analysis of airflow (AF) recordings to help in sleep apnea-hypopnea syndrome (SAHS) diagnosis. The objective is to estimate the apnea-hypopnea index (AHI) by means of spectral features from AF data. Multiple linear regression (MLR) was used for this purpose. A training group is used to obtain two MLR models: the first one consisting of features obtained from the full PSDs (MLR_{full}) and the second one consisting of features from a new frequency band of interest (MLR_{band}). Then a test group is used to validate the final model. The correlation of spectral features and MLR models with AHI was compared by means of Pearson's coefficient (ρ). MLR_{band} reached the highest ρ (0.809). Four different AHI decision thresholds were used to evaluate MLR_{band} ability to distinguish the severity of SAHS. The accuracy achieved was higher as the threshold increased (69.7%, 75.3%, 80.9%, 87.6%) These results suggest that the automated estimation of AHI through spectral features can provide useful knowledge about SAHS severity.**

I. INTRODUCTION

The sleep apnea hypopnea syndrome (SAHS) is a prevalent disease characterized by recurrent events of complete cessation (apneas) and significant reduction (hypopneas) of breathing during sleep [1]. SAHS has been associated with other diseases such as cardiac failure, stroke, atrial fibrillation and sudden cardiac death [2]. Moreover, daytime sleepiness caused by SAHS has been recognised as a risk factor for occupational accidents and motor-vehicle collisions [3], [4].

The standard test for SAHS diagnosis is polysomnography (PSG) [5]. PSG is a complex test since many physiological signals are recorded from patients while they are asleep. Electrocardiogram (ECG), electroencephalogram (EEG), airflow (AF) and oxygen saturation (SpO_2) are some examples [6]. Furthermore, PSG needs overnight supervision of patients and offline visual inspection of signals in order to derive the apnea-hypopnea index (AHI). AHI estimates the number of apneas and hypopneas events per hour during sleep and is used to evaluate SAHS severity [1]. PSG is associated to long waiting lists and increased delay time for a final diagnosis [7]. Thus, simpler methods to detect SAHS are needed. One common alternative is to analyze a reduced set of signals from overnight PSG [8].

This paper focuses on the analysis of single channel AF. There exist several previous studies focused on SAHS detection by means of AF analysis [9]-[12]. Most of them are aimed to derive a respiratory disturbance index (RDI) by counting events of reduction in the AF amplitude. However, despite its usefulness for research purposes, RDI is not designated as the reference diagnostic parameter [6].

The main objective of the study is to automatically derive the AHI from AF. The extraction of features from power spectral density (PSD) of recordings is proposed for this purpose. This analysis has proved to be useful in SAHS detection [8], [10], [13]. Then multiple linear regression (MLR) technique is applied to spectral features in order to estimate AHI. This estimation could provide useful knowledge about SAHS severity. MLR is a common statistic technique for regression analysis [14]. It has been used in previous studies for AHI estimation from SpO_2 data [15], [16]. Finally, the ability of the proposed methodology to distinguish SAHS severity is assessed. Four severity thresholds of the AHI are used for this purpose.

II. SUBJECTS AND SIGNALS

The study involved 148 adult subjects, 79.1% men and 20.9% women. All of them were suspected of suffering from SAHS due to previous symptoms such as daytime sleepiness, loud snoring and nocturnal choking and awakenings. Subjects underwent overnight PSG in the sleep unit of the Hospital Universitario Rio Hortega in Valladolid, Spain. The Review Board on Human Studies accepted the protocol, and all subjects gave their informed consent to participate in the study. Physicians established the AHI threshold for a positive diagnosis in 10 events per hour (e/h). Accordingly, 100 subjects were considered SAHS-positive and 48 SAHS-negative. Apnea was defined as the cessation of AF for 10 seconds at least. Hypopnea was defined as a 30% reduction in AF during 10 seconds or more, accompanied by a 4% or more decrease in the saturation of haemoglobin. The subjects were randomly divided into two groups: training group (40%) and test group (60%).

AF signals were obtained from PSG (Alice 5, Respironics, Philips Healthcare, The Netherlands). A thermistor (Pro-Tech) was used to measure AF at sample rate of 10 Hz. The length of the recordings was 7.24 ± 0.38 h

*This research was supported in part by the "Ministerio de Economía y Competitividad" under project TEC2011-22987, the Proyecto Cero 2011on Ageing from Fundación General CSIC, the project VA111A11-2 and the PIRTU contract (Orden EDU/1204/2010) from Consejería de Educación (Junta de Castilla y León), and the European Social Found.

G. C. Gutiérrez-Tobal, R. Hornero, D. Álvarez, J. V. Marcos and C. Gómez are with the Biomedical Engineering Group of the University of Valladolid, Valladolid, Spain (corresponding author phone: 983423000 ext.: 3981 e-mail: cargom@tel.uva.es)

F. del Campo is with the sleep unit of Hospital Universitario Rio Hortega in Valladolid, Spain (e-mail: fsas@telefonica.net).

(mean ± sd). Table I summarizes demographic and clinical data from the population under study.

III. METHODS

The PSD of AF recordings was estimated by means of the non-parametric Welch method. This method is suitable when analyzing non-stationary signals [17]. Each recording was divided into Hamming windows of 2048 samples, with 50% overlap and 4096-points DFTs, in order to compute the final estimation.

A. Definition of bands of interest

The frequency band of interest was defined as the spectral region in which the statistically significant differences between the PSD of both populations (SAHS-positive and SAHS-negative) were higher. Thus, the non-parametric Kruskal-Wallis test was used to compute the *p*-value along frequencies. A band of interest was derived (0.024-0.056 Hz.) by choosing the range of frequencies for which the *p*-value was the lowest in the training group. Fig. 1 displays a joint representation of the PSD of each population. The frequency band of interest is also shown.

B. Spectral features

Several spectral features were extracted from the PSDs of AF recordings. All of them were obtained from the band of interest as well as the full PSD. First-to-fourth statistical moments (Mf_1-Mf_4), peak amplitude (PA) and Wootters distance (WD) were used to reflect the recurrence of apnoeic events. The arithmetic mean (Mf_1), standard deviation (Mf_2), skewness (Mf_3) and kurtosis (Mf_4) quantify the central tendency, dispersion, asymmetry and peakedness of PSD values, respectively. PA is the maximum of the PSD in a given frequency region and can be computed as follows:

$$PA = \max_{f}\{PSD(f)\},\ f(Hz) \in [f_i, f_N], i = 1,2,...,N\ , \tag{1}$$

where N is the number of points in the band and f_i each frequency component. Finally, WD is a disequilibrium measurement [18]. It reaches higher values if the spectrum is condensed into a narrow frequency band, whereas lower values are obtained if spectral components are distributed along frequencies:

$$WD = arc\cos\{\sum_{f_i=f_1}^{f_2} \sqrt{PSD_n(f_i)}\ \sqrt{1/N}\}\ , \tag{2}$$

where PSD_n is the normalized PSD.

C. Multiple linear regression

Given the set of features $x_1, x_2,...,x_k$ extracted from AF data, the aim is to approximate the functional relationship between them and the AHI. Regression analysis using MLR was used for this purpose, being AHI the target variable. The following functional form is assumed for the approximation [14]:

$$y = w_0 + w_1 x_1 + ... + w_k x_k = \mathbf{w}^T \mathbf{x}\ , \tag{3}$$

where $\mathbf{w} = (w_0, w_1,\ldots,w_k)^T$ are unknown constant parameters computed according to sum-of-squares error minimization [14]. Thus, the MLR model assumes a linear relationship between the explanatory variables and the dependent one.

In this study, **w** was computed from the training group. Then **w** was applied to the test group to validate the diagnostic ability of the model. Two MLR models were obtained: MLR_{full} was computed by using all the spectral features from the full PSD and MLR_{band} was computed by the use of all the spectral features from the new band of interest.

D. Statistical analysis

The Pearson's correlation coefficient ρ was used to evaluate the linear relationships between the AHI from PSG and the estimated AHI. It was also used to assess the linear correlation between the spectral features and the true AHI. A ρ value close to zero indicates poor linear relationship and ρ close to ±1 indicates high linear relationship. Sensitivity (percentage of actual positives correctly identified), specificity (percentage of actual negatives correctly identified) and accuracy (proportion of correct classifications) were used to assess the performance of the

TABLE I. DEMOGRAPHIC AND CLINICAL DATA

	All subjects	Training group	Test group
Subjects (n)	148	59	89
Age (years)	50.9 ± 11.7	49.2 ± 11.3	51.9 ± 11.8
BMI (kg/m^2)	29.2 ± 4.7	28.3 ± 4.1	29.8 ± 5.0
AHI	23.5 ± 24.1	19.1 ± 17.5	26.5 ± 27.4

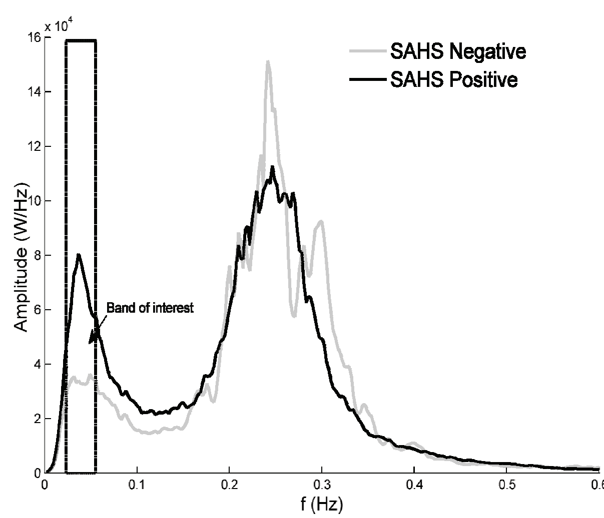

Figure 1. Joint representation of the PSD of each population in the training group.

AHI estimation when it is used to differentiate subjects among several degrees of SAHS severity. The agreement between the estimated AHI and the AHI from PSG was evaluated by means of a Bland-Altman plot.

IV. RESULTS

Table II displays the values of Pearson's correlation coefficient ρ between spectral features from the test group and the corresponding AHI values from PSG. Most of features from the band of interest reached higher values of ρ than the features extracted from the full PSD. The poorest value of ρ was reached by Mf_2 (-0.009) whereas the highest value was reached by Mf_1^a (0.672).

Table III shows the Pearson's correlation coefficient between true AHI and the AHI from estimators on the test group. MLR_{band} clearly improves (0.809) the values of ρ reached by each of the features as well as by MLR_{full}.

Fig. 2 shows the Bland-Altman plot comparing the estimated AHI for MLR_{band} and the AHI from PSG (mean vs. difference). The mean of the difference (bias) reached a negative value (-1.18 e/h) near zero. The scatter plot indicates overestimation for small values of AHI and underestimation when the AHI is higher. The 95% limits of agreement (mean of the difference ± 1.96*SD of the difference) reached 30.71 e/h and -33.07 e/h.

Table IV summarizes the results of the MLR_{band} model to distinguish the severity of SAHS. Four AHI cut-offs have been used (5, 10, 15 and 30 e/h). These correspond to common clinical thresholds. The accuracy increases as a more severity degree of SAHS is considered. The highest accuracy is achieved for a cut-off of 30 e/h (87.6%). The accuracy reaches 75.3% and 80.9% for 10 e/h and 15 e/h, respectively. The lowest value was performed for a 5 e/h threshold (69.7%).

TABLE II. PEARSON'S CORRELATION COEFFICENTS BETWEEN SPECTRAL FEATURES FROM TEST GROUP AND AHI FROM PSG

Feature	ρ	Feature	ρ
Mf_1	0.061	Mf_1^a	0.672
Mf_2	-0.009	Mf_2^a	0.560
Mf_3	-0.118	Mf_3^a	0.293
Mf_4	-0.015	Mf_4^a	-0.082
PA	-0.010	PA^a	0.632
WD	-0.294	WD^a	0.669

a. Features extracted from the frequency band of interest

TABLE III. PEARSON'S CORRELATION COEFFICENTS BETWEEN MLR ESTIMATIONS FROM TEST GROUP AND AHI FROM PSG

AHI estimation	ρ
MLR_{full}	0.381
MLR_{band}	0.809

TABLE IV. RESULTS FROM MLR_{BAND} MODEL

AHI thres. (e/h)	5	10	15	30
Sensitivity(%)	100.0	95.0	93.5	76.7
Specificity(%)	0.0	34.5	67.4	93.2
Accuracy(%)	69.7	75.3	80.9	87.6

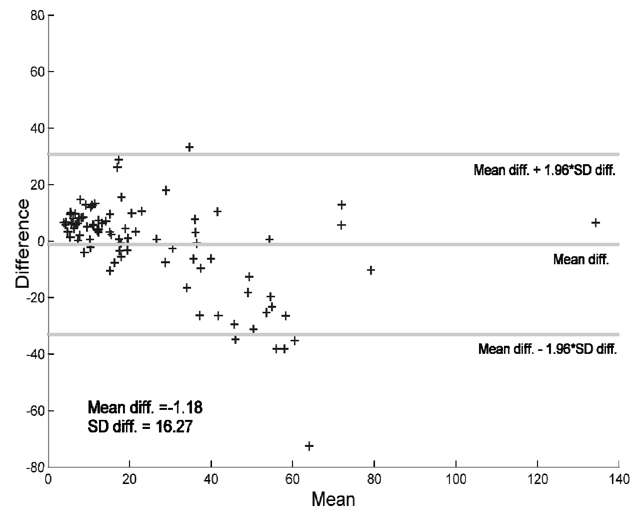

Figure 2. Bland-Altman plot comparing the estimated AHI and the AHI from the PSG.

V. DISCUSSION AND CONCLUSION

The AHI was estimated applying MLR to spectral features from AF recordings. Pearson's coefficient ρ was used to assess correlation between the spectral features, AHI from MLR models and AHI from PSG. The features extracted from a frequency band of interest (0.024-0.056 Hz.) reached higher ρ than the features obtained from the full PSDs. Moreover, MLR_{band} reached the highest ρ (0.809). Both results suggest that this new frequency band of interest is highly related with apneic events. Furthermore, this band is consistent with pathophysiology since apneic events must last 10 seconds or more [5], i.e. their frequency must be located under 0.1 Hz.

The agreement between the estimated AHI (MLR_{band}) and the true AHI was evaluated by means of Bland-Altman plot. Results show an overestimation tendency as the mean is low and an underestimation tendency as the mean is higher. The behaviour of the Bland-Altman plot is reflected by the results from the assessment of estimated AHI ability to discriminate degrees of SAHS. Thus, the specificity reaches 0% when the AHI threshold equals 5 e/h, resulting in poor accuracy (69.7%). However, accuracy increases (75.3%, 80.9%, 87.6%) as the threshold is higher (10, 15, 30 e/h). This reflects an elevated ability to distinguish high degrees of SAHS severity.

Several works are focused on AHI estimation by analyzing different data sets. El-Solh *et al*. derived AHI from clinical and demographic data by using a multilayer perceptron neural network (MLP) [19]. Magalang *et al*. computed AHI by the use of oxymetric indices as inputs to a multivariate adaptive regression splines (MARS) model [20]. Roche *et al.* obtained a MLR model combining both clinical and oximetry features [16]. In a previous study of our research group, Marcos *et al*. predicted AHI applying MLR and MLP to 14 features extracted from SpO_2 recordings [15]. For an AHI threshold = 15 e/h, the results from these studies range 61%-95% sensitivity, 60%-90% specificity and 62%-93% accuracy.

Data obtained from SpO_2 recordings indirectly shows the nature of apnoeic events. Nevertheless, AF waveform is directly modified by apneas and hipopneas. No studies have been found focused on AHI estimation through features from AF. However, some are focused on counting respiratory events and estimating RDI [9]-[12]. The performance of these works range 81%-86% sensitivity, 83%-90% specificity and 0.73-0.95 Pearson's correlation coefficient, for an AHI cut-off =15 e/h.

Some limitations related with this work have to be pointed out. The number of subjects under study should be larger to assess the generality of results. Moreover, all subjects were suspected of having SAHS before PSG test. Non-suspected people should be added to the study in order to evaluate general applicability of the methodology. Another limitation is related with the use of a thermistor instead of a nasal prong pressure (NPP) sensor to acquire AF. Although the American Academy of Sleep Medicine (AASM) recommends the use of both types of sensors [6], it is known that NPP improves the performance of thermistor to detect obstructive respiratory events [21]. On the other hand, a different choice of the *p*-value threshold when computing the band of interest can change its length and, consequently, the final performance of the proposed method. Finally, the MLR used to estimate AHI assumes linear relationships between the explanatory variables and the dependent one. Therefore, other methods for the estimation of AHI should be used to assess the non-linear relationships with AF features. The use of artificial neural networks is a future goal for this purpose.

In summary, a new frequency band of interest related to apneic events was defined. Spectral features from this band were used to obtain a MLR model and estimate AHI. The ability of the estimated AHI to distinguish SAHS severity was assessed by the use of different AHI cut-offs. For the thresholds AHI = 15 e/h and AHI = 30 e/h, accuracies over 80% were achieved (80.9% and 87.6%, respectively). Therefore, the automated estimation of AHI through spectral features from AF recordings can provide useful knowledge about SAHS severity.

REFERENCES

[1] S. P. Patil, H. Schneider, A. R. Schwartz, and P. L. Smith, "Adult Obstructive Apnea," *Chest*, vol. 132(1), pp. 325-337, 2007.

[2] F. Lopez-Jiménez, F. H. S. Kuniyoshi, A. Gami, and V. K. Somers, "Obstructive Sleep Apnea," *Chest*, vol. 133(1), pp 793-804, 2008.

[3] E. Lindberg, N. Carter, T. Gislason, and C. Janson, "Role of Snoring and Daytime Sleepiness in Occupational Accidents," *American Journal of Respiratory and Critical Care Medicine*, vol. 164, pp. 2031-2035, 2001.

[4] A. Sassani, L. J. Findley, M. Kryger, E. Goldlust, C. George, and T. M. Davidson, "Reducing Motor-Vehicle Colissions, Costs, and Fatalities by Treating Obstructive Sleep Apnea Syndrome" *Sleep*, vol. 27(3), pp. 453-458, 2003.

[5] W. W. Flemons, M. R. Littner, J. A. Rowley, P. Gay, W. M. Anderson, D. W. Hudgel, R. D. McEvoy, and D. I. Loube, "Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature ," *Chest*, vol. 124, pp. 1543-1579, 2003.

[6] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan. *The AASM Manual for the Scoring of Sleep and Associated Events*. Manual, American Academy of Sleep Medicine, 2007.

[7] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea," *American Journal of Respiratory and Critical Care Medicine*, vol. 169, pp. 668-672, 2004.

[8] T. Penzel, J. McNames, P. de Chazal, B. Raymond, A. Murray, and G. Moody, "Systematic Comparison of Different Algorithms for Apnoea Detection Based on Electrocardiogram Recordings," *Medical and Biological Engineering Computing*, vol. 40, pp. 402-407, 2002.

[9] T. Shochat, N. Hadas, M. Kerkhofs, A. Herchuelz, T. Penzel, J. H. Peter, and P. Lavie, "The SleepStripTM: An Apnoea Screener for the Early Detection of Sleep Apnoea Syndrome," *European Respiratory Journal*, vol. 19, pp. 121-126, 2002.

[10] H. Nakano, T. Tanigawa, T. Furukawa, and S. Nishima, "Automatic Detection of Sleep-disordered breathing from a single-channel airflow record," *European Respiratory Journal*, vol. 29, pp. 728-736, 2007.

[11] A. C. Tonelli, D. Martinez, L.F. T. Vasconcelos, S. C. Gonçalves, M. C. Lenz, S. Costa, M. Gus, E. Oliveira, L. Beltrami, and F. Danni, "Diagnosis of Obstructive Sleep Apnea Syndrome and Its Outcomes With Home Portable Monitoring," *Chest*, vol. 135, pp. 330-336, 2009.

[12] L. M. Rofail, K. K. H. Wong, G. Unger, G. B. Marks, and R. R. Grunstein, "The role of single-channel nasal airflow pressure transducer in the diagnosis of OSA in the sleep laboratory,". *Journal of Clinical Sleep Medicine*, vol. 6(4), pp. 349-356, 2010.

[13] D. Álvarez, R. Hornero, J. V. Marcos, and F. Del Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2816–2824, 2010.

[14] J. D. Jobson, *Applied Multivariate Data Analysis. Volume I: Regression and Experimental Design*. New York: Springer, 1991.

[15] J. V. Marcos, R. Hornero, D. Álvarez, M. Aboy, and F. Del Campo, "Automated prediction of the apnea-hypopnea index from nocturnal oximetry recordings," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 141–149, 2012.

[16] N. Roche, B. Herer, C. Roig, and G. Huchon, "Prospective testing of two models based on clinical and oximetric variables for prediction of obstructive sleep apnea," *Chest*, vol. 121, pp. 747–752, 2002.

[17] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on time Averaging Over Short, Modified Periodograms," *IEEE Transactions on Audio Electroacoustics*, vol. AU-15 pp. 70-73, 1967.

[18] W. K. Wootters, "Statistical distance and Hilbert space," *Physical Review D*, vol. 23(2), pp. 357-362, 1981.

[19] A. A. El-Solh, M. J. Mador, E. Ten-Brock, D. W. Shucard, M. Abul-Khoudoud, and B. J. B. Grant, "Validity of neural network in sleep apnea," *Sleep*, vol. 22, pp. 105–111, 1999.

[20] U. J. Magalang, J. Dmochowski, S. Veeramachaneni, A. Draw, M. J. Mador, A. El-Solh, and B. J. B. Grant, "Prediction of the apneahypopneaindex from overnight pulse oximetry," *Chest*, vol. 124,pp. 1694–1701, 2003.

[21] A. BaHammamm, "Comparison of nasal prong pressure and thermistors measurements for detecting respiratory events during sleep," *Respiration*, vol. 71, pp. 385-390, 2004.