Exploring the Spectral Information of Airflow Recordings to Help in Pediatric Obstructive Sleep Apnea-Hypopnea Syndrome Diagnosis

Gonzalo C. Gutiérrez-Tobal, *Student Member*, *IEEE*, Daniel Álvarez, *Member, IEEE*, M. Luz Alonso, Joaquín Terán, Félix del Campo, and Roberto Hornero*, *Senior Member*, *IEEE*

Abstract— **This work aims at studying the usefulness of the spectral information contained in airflow (AF) recordings in the context of Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) in children. To achieve this goal, we defined two spectral bands of interest related to the occurrence of apneas and hypopneas. We characterized these bands by extracting six common spectral features from each one. Two out of the 12 features reached higher diagnostic ability than the 3% oxygen desaturation index (ODI3), a clinical parameter commonly used as screener for OSAHS. Additionally, the stepwise logistic regression (SLR) feature-selection algorithm showed that the information contained in the two bands was complementary, both between them and with ODI3. Finally, the logistic regression method involving spectral features from the two bands, as well as ODI3, achieved high diagnostic performance after a bootstrap validation procedure (84.6±9.6 sensitivity, 87.2±9.1 specificity, 85.8±5.2 accuracy, and 0.969±0.03 area under ROC curve). These results suggest that the spectral information from AF is helpful to detect OSAHS in children.**

I. INTRODUCTION

The Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) is characterized by recurrent episodes of apnea (complete absence of airflow) and hypopnea (significant reduction of airflow) during sleep [1]. Apneic events lead to oxygen desaturations and arousals which prevent patients from resting while sleeping, disturbing both their health and quality of life [1]. Common symptomatology of pediatric OSAHS includes overnight snoring and sleep difficulties [2], which derive in other daytime symptoms and illnesses such as cognitive and behavioral irregularities, abnormal growth, and cardiovascular malfunctions [3]. Up to 6 % of children affected has been reported [3], indicating the high prevalence of OSAHS.

OSAHS is diagnosed by means of nocturnal polysomnography (PSG) test, which acts as the "gold standard" [1]. PSG requires recording a wide range of physiological signals from patients overnight, including electroencephalogram (EEG), electrocardiogram (ECG), oxygen saturation (SpO_2), and airflow (AF) [1]. Hence, the necessary equipment is both complex and costly [4]. OSAHS and its severity are established according to the apnea hypopnea index (AHI), which estimates the number of apneic events per hour. To derive AHI, physicians need to examine all the physiological recordings. Consequently, PSG is also time-consuming [4]. Furthermore, PSG may interfere with the sleep routine of patients since it has to be carried out into a specialized sleep unit.

To overcome these drawbacks a number of alternatives have been studied. One common approach is the computation of different estimations of AHI, such as the respiratory disturbance index (RDI) or the oxygen desaturation index (ODI), using a reduced set of signals from PSG [5]. Alternatively, single-channel signals such as SpO_2 and photoplethysmography (PPG) have been recently assessed to automatically detect OSAHS, both in adults and children [6-9].

In this paper, the analysis of single-channel AF recordings from children is proposed. The main objective is to evaluate eventual differences between OSAHS patients (OSAHS-positive) and no-OSAHS subjects (OSAHS-negative). The study of AF is a straightforward choice since apneas and hypopneas are defined on the basis of its amplitude variations [10]. The recurrence of apneic events naturally leads to a study in the frequency domain. In this sense, recent works have shown that OSAHS modifies the spectral content of AF recordings from adults at certain frequencies, and that the information contained in such frequencies is useful in OSAHS detection [11, 12]. However, no studies have been found applying a similar analysis to AF recordings from children. According to the above mentioned, we pose two research questions:

i. How OSAHS alters the spectral information of AF recordings from children?

ii. Are these alterations useful to distinguish OSAHS in children?

To answer the first one we perform an exploratory analysis of the power spectral density (PSD) of the AF recordings. We look for spectral bands of interest, as well as their characterization. To assess the second question, we evaluate the diagnostic performance of several common spectral features, separately and in a jointly way through logistic regression models. Additionally, we also

G. C. Gutiérrez-Tobal, D. Álvarez, and R. Hornero, are with the Biomedical Engineering Group (University of Valladolid, Spain) (corresponding author e-mail: robhor@tel.uva.es)

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

M. Luz Alonso and Joaquín Terán are with the sleep unit of the Hospital Universitario de Burgos, Spain (e-mail: mlalonso@hubu.es).

assess the diagnostic performance of 3% ODI (ODI3) for comparison purposes. Our hypothesis is that the spectral information contained in single-channel AF could be useful to help in the diagnosis of pediatric OSAHS.

II. SUBJECTS AND SIGNALS UNDER STUDY

This study involved AF recordings from 50 children ranging 3 to 13 years old (24 OSAHS-negative and 26 OSAHS-positive). All of them were remitted to the unit of respiratory sleep disorders of the Hospital Universitario de Burgos (Spain), due to clinical suspicion of OSAHS. AF was acquired during a polygraphy test performed at patients' home. The sensor used to obtain AF was a thermistor and the sample rate was 100 Hz. The physicians used the AHI derived from PSG to establish OSAHS. As stated by the American Academy of Sleep Medicine (AASM), they scored apneas and hypopneas as events lasting 2 missed respiratory cycles, at least [10]. An AHI threshold of 3 events/hour was used to distinguish OSAHS-positive from OSAHS-negative subjects. The Review Board on Human Studies accepted the protocol and an informed consent was obtained for each subject. Table I shows clinical and demographical data from the subjects under study. No statistical significant differences in body mass index (BMI) or age were found between groups (p-value $\gg 0.01$).

III. METHODS

A. Power Spectral Density analysis

We computed the PSD of each AF recording to explore eventual differences between the spectral information of OSAHS-positive and OSAHS-negative groups. The estimation of the PSDs was carried out by the non-parametric Welch method, which is suitable for non-stationary signals [13]. A Hamming window of 2^{15} samples (50% overlap) along with a discrete Fourier transform of 2^{16} samples were used. Each PSD was normalized (PSDn) dividing all their spectral components by their corresponding total power.

To define the spectral bands of interest, we looked for statistical significant differences between PSDns from OSAHS-positive and OSAHS-negative groups. We used a p-value based methodology consisting in applying the Mann-Whitney U test to the amplitude values of the PSDns from both groups, at each frequency [11, 12]. Fig. 1 shows the median value of the PSDns from OSAHS-positive (red) and OSAHS-negative (blue) samples at each frequency. It also shows the p-value obtained in the comparison of both groups (grey). We selected those spectral bands in which the p-value were lower than 0.01 at least in the 90% of their components. Two bands were defined: 0.119-0.192 Hz (BW1); 0.784-0.890 Hz (BW2). These bands were characterized in every subject by extracting six common spectral features from each one:

- Maximum and minimum amplitude (MA, mA), computed as the highest and the lowest PSDn values in each band.

- First to fourth statistical moments (Mf_1-Mf_4), i.e. mean, standard deviation, skewness, and kurtosis, respectively.

TABLE I. DEMOGRAPHIC AND CLINICAL DATA

Features	All	OSAHS+	OSAHS-
# Subjects	50	26	24
Age* (years)	5.3±2.5	5.4±2.7	5.2±2.4
Male (%)	46.0	38.5	54.2
BMI+ (kg/m²)	16.5±2.5	16.9±3.0	16.1±1.7
Rcording Time (h)	8.9±0.8	8.8±1.0	9.0±0.5
AHI (e/h)	9.9±13.8	17.9±15.4	1.3±0.8

BMI: Body Mass Index; AHI: Apnea Hypopnea Index; *p-value = 0.76; +p-value = 0.94

B. Logistic Regression: feature selection and classification

The logistic regression (LR) method is a supervised learning algorithm which estimates the posterior probability of a given instance $\mathbf{x}_i$ belonging certain class C_k (in our case, C_k = OSAHS-positive or OSAHS-negative). Hence, the posterior probability $p(C_k; \mathbf{x}_i)$ is computed through the logistic function:

$$p(C_k \mid \mathbf{x}_i) = \frac{e^{\beta_0 + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}_i}}{1 + e^{\beta_0 + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}_i}}, \qquad (1)$$

where β_0 and $\boldsymbol{\beta}$ are obtained by the weighted least squares minimization procedure [14]. Thus, an instance $\mathbf{x}_i$ is assigned to the class with larger posterior probability.

We used LR with two purposes. First, it was used to automatically select those relevant and non-redundant features. This was performed through the stepwise LR method (SLR), proposed by Hosmer and Lemeshow [14]. Specifically, we applied the forward-selection backward-elimination algorithm. Second, LR was also used to assess the joint potentiality of the selected features from BW1 and BW2 to predict OSAHS in children.

C. Statistical analysis and results validation

The Mann-Whitney U test was used to evaluate statistical differences in the spectral features from OSAHS-positive and OSAHS-negative groups. Sensitivity (Se, percentage of OSAHS-positive subjects rightly classified), specificity (Sp, percentage of OSAHS-negative subjects rightly classified), and accuracy (Acc, overall percentage of subjects rightly classified) were used to measure the diagnostic ability of both single features and LR models. To find an optimum threshold, u_o, for the assessment of each single feature, a receiver operating-characteristics (ROC) analysis was done. The area under the ROC (AROC) was also computed in each case.

We used the *bootstrap 0.632* algorithm to validate our results [15]. Thus, since the number of subjects in our database is $n=50$, 50 new groups of size 50 were built by resampling with replacement from the original one. These acted as training groups. For each training group a number of elements are repeated. The instances not included in them acted as test groups. Following this procedure, a measure m obtained from the test set would be a pessimistic estimation of the true one [15]. Hence, both the training and the test groups are used to compute m by
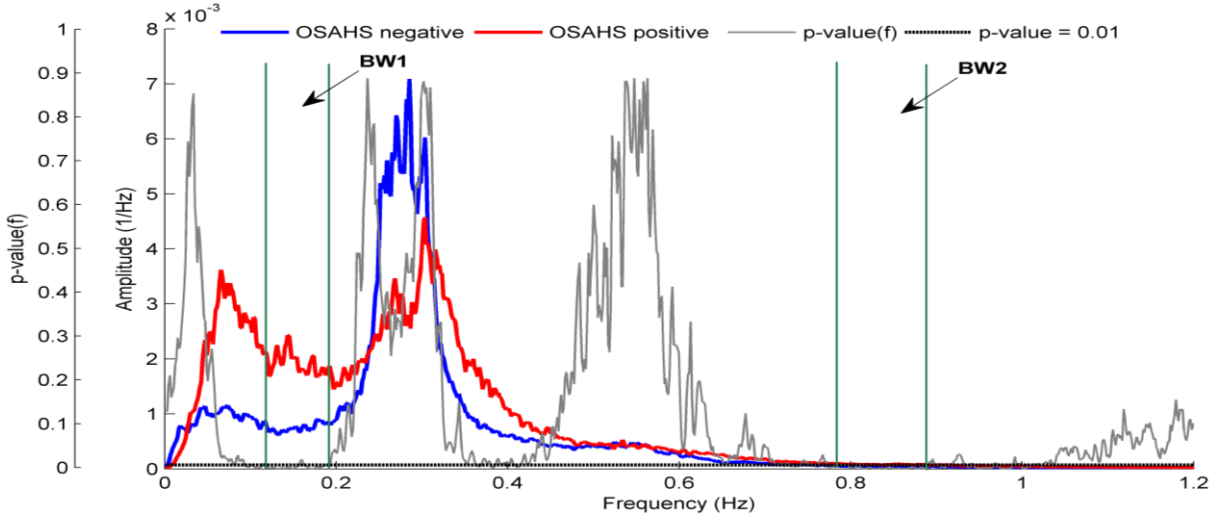
Figure 1. Bands of interest according to the spectral analysis

weighting their corresponding estimations as follows [15]:

$$m = 0.632 \cdot m_{test} + 0.368 \cdot m_{training} \quad . \qquad (2)$$

Finally, the 50 estimations of m are averaged to show a global performance.

IV. RESULTS

A. Descriptive analysis and feature selection

Table II summarizes the averaged values (mean ± standard deviation) of each spectral feature. Consistent with Fig. 1, the values of MA, mA, and Mf_1 in BW1 and BW2 were significantly higher in OSAHS-positive than in OSAHS-negative subjects. Near to significant differences were found in Mf_2 from both bands, and there were no differences in Mf_3 and Mf_4.

SLR was used twice to select relevant and non-redundant features. First, we applied SLR to the 12 spectral features previously obtained. Thus, mA from BW1, and Mf_3 and Mf_4 from BW2 were automatically selected by SLR (SLR_{Spec}). Then, we added ODI3 to the selection process. In this case, ODI3, mA and Mf_4 from BW1, as well as Mf_3 from BW2 were automatically chosen ($SLR_{Spec-ODI3}$).

B. Diagnostic performance

Table III shows the diagnostic performance of the spectral features after the *bootstrap 0.632* procedure. Se, Sp, and Acc values (mean ± standard deviation) were obtained by weighting the training and test estimations as stated by (2), and averaging the results from the 50 training-test group pairs. The values of u_o and AROC are meaningful as parameters from training groups. Hence, only these were used to estimate them. Among the spectral features, the highest Acc and AROC were reached by mA from BW1 (76.5 ± 3.8 %; 0.744 ± 0.07, respectively).

Table IV includes the diagnostic performance of ODI3 from oximeter as well as that from the SLR_{Spec} and the $SLR_{Spec-ODI3}$ models. In the case of the LR models, a 0.5 posterior probability was maintained as the typical threshold to decide between OSAHS-positive and OSAHS-negative classes. The performance of ODI3 did not improve that of mA from BW1 and BW2. Additionally, SLR_{Spec}, which only uses spectral information, outperformed all the single features in terms of Acc and AROC. Finally, the highest results were obtained by $SLR_{Spec-ODI3}$, combining spectral information from AF with ODI3 from SpO_2 (84.6 ± 9.6 % Se, 87.2 ± 9.1 % Sp, 85.8 ± 5.2 % Acc, and 0.969 ± 0.03 AROC).

V. DISCUSSION AND CONCLUSIONS

A spectral analysis of AF recordings from children was performed in the context of OSAHS. We found that the spectral power of AF was significantly higher in OSAHS-positive subjects at two frequency bands below (BW1) and above (BW2) the typical respiratory range in children (0.22-0.43 Hz., aprox.) [16]. BW1 (0.119-0.192 Hz.) is consistent with the occurrence of apneas and hypopneas since these were scored when lasted 2 missed breaths, at least [10]. Differences in BW2 (0.784-0.890 Hz.) could be explained by the typical respiratory overexertion after an apneic event, which increases the respiratory rate [5].

Useful features were obtained from BW1 and BW2. Thus, 6 out of 12 were significantly different in OSAHS-positive than in OSAHS-negative subjects. Additionally,

TABLE II. FEATURE AVERAGED VALUES FOR SAHS-POSITIVE AND SAHS-NEGATIVE GROUPS

	BW1			BW2		
	mean ± sd			mean ± sd		
	OSAHS +	OSAHS -	p	OSAHS +	OSAHS -	p
MA (10^{-3})	4.1±5.7	1.9±2.0	**<0.01**	0.21±0.23	0.14±0.21	**<0.01**
mA (10^{-4})	13.0±8.0	5.8±2.9	**<<0.01**	0.87±0.78	0.38±0.28	**<<0.01**
Mf_1 (10^{-3})	2.2±1.7	1.1±0.8	**<0.01**	0.14±0.13	0.07±0.07	**<0.01**
Mf_2 (10^{-4})	7.5±16.6	3.2±4.3	0.045	0.28±0.31	0.25±0.48	0.029
Mf_3 (10^{-1})	6.5±0.5	6.5±0.5	0.993	4.7±4.8	5.7±4.6	0.541
Mf_4 (10^{0})	3.0±1.0	3.1±1.0	0.749	3.1±1.2	2.8±0.9	0.356

p: p-value of the Mann-Whitney U test; sd: standard deviation.

TABLE III. Diagnostic performance of the spectral features

BW1	u_o	Se (%)	Sp (%)	Acc (%)	AROC
MA	$2.4 \cdot 10^{-3}$	58.8±12.7	74.0±14.7	66.0±5.8	0.636±0.08
mA	$8.0 \cdot 10^{-4}$	72.9±8.5	81.0±8.9	**76.5±3.8**	**0.744±0.07**
Mf$_1$	$1.3 \cdot 10^{-3}$	67.3±11.2	73.4±10.2	69.8±4.7	0.668±0.08
Mf$_2$	$2.8 \cdot 10^{-4}$	60.9±11.8	65.8±14.3	63.1±5.1	0.581±0.10
Mf$_3$	$6.1 \cdot 10^{-1}$	52.2±9.8	55.5±13.1	53.9±6.1	0.533±0.08
Mf$_4$	$2.7 \cdot 10^{0}$	50.4±12.2	56.1±12.2	52.8±5.0	0.545±8.5
BW2					
MA	$8.8 \cdot 10^{-5}$	79.4±10.3	63.8±10.3	71.9±4.7	0.672±0.07
mA	$3.4 \cdot 10^{-5}$	85.5±12.4	65.8±8.2	**75.9±5.5**	**0.729±0.07**
Mf$_1$	$5.6 \cdot 10^{-5}$	83.3±8.9	64.6±8.7	74.3±4.5	0.690±0.09
Mf$_2$	$1.1 \cdot 10^{-5}$	76.4±10.3	65.1±9.2	70.9±5.7	0.617±0.09
Mf$_3$	$4.5 \cdot 10^{-1}$	52.1±12.7	50.5±13.3	51.3±6.2	0.541±0.08
Mf$_4$	$2.7 \cdot 10^{0}$	57.7±12.4	60.5±9.0	59.1±5.0	0.563±0.07

TABLE IV. Diagnostic performance of ODI3 and the logistic regression models

	u_o	Se (%)	Sp (%)	Acc (%)	AROC
ODI3	2.6	72.2±12.3	79.7±15.5	75.8±3.5	0.702±0.08
SLR$_{Spec}$	--	**78.5±11.8**	**80.6±9.9**	**78.9±4.9**	**0.902±0.04**
SLR$_{Spec-ODI3}$	--	**84.6±9.6**	**87.2±9.1**	**85.8±5.2**	**0.969±0.03**

mA from BW1 and BW2 outperformed the ODI3, both in Acc and AROC. The study showed that the information from the two bands is complementary since SLR selected features from the two of them to build the SLR$_{Spec}$ and SLR$_{Spec-ODI3}$ models. Complementarity between the information from BW1-BW2 and ODI3 was also showed.

Both SLR$_{Spec}$ and SLR$_{Spec-ODI3}$ outperformed all the single features, including ODI3. Particularly high was the diagnostic performance of SLR$_{Spec-ODI3}$, which widely improved the performance of a 6-channel respiratory polygraphy (RP) at each statistic (81.5 % Se, 80.8 % Sp, 81.1 % Acc, and 0.876 AROC) [5], only requiring information from 2 channels (thermistor and oximeter). SLR$_{Spec}$ also improved RP in terms of AROC, and performed similarly in Se, Sp, and Acc. Other recent studies also analyzed physiological signals to help in OSAHS diagnosis in children. Shouldice *et al.* used 50 ECG recordings, and reached 85.7 % Se, 81.8 % Sp, and 84 % Acc by applying a quadratic linear discriminant to 23 features [17]. Gil *et al.* investigated the diagnostic usefulness of the information contained in 21 PPG time series, reaching 75.0 % Se, 85.7% Sp, and 80.0% Acc [7]. Finally, Garde *et al.* achieved 80.0 % Se, 92.1 % Sp, and 86.8% Acc applying a linear discriminant methodology to five features, from time and frequency domain, obtained from 68 SpO$_2$ recordings [9].

Some limitations have to be addressed in this study. The sample size should be larger to empower the generalization ability of our results. A larger sample would let us define the bands of interest through an independent set of subjects. Thus, our spectral bands were defined from our specific database. However, they were consistent with the pathophysiology of the apneic events. A wide sample of subjects would be also useful to optimize the set of selected features. Future application of different spectral or non-linear measures could complement our findings.

To the best of our knowledge, this is the first time that the spectral information of AF recordings from children is analyzed in the context of OSAHS. We showed that OSAHS in children modifies the PSD of AF at two abnormal respiratory bands. Single features from these bands outperformed the diagnostic ability of ODI3. Additionally, the information contained in the two bands showed complementarity both between them and with ODI3. An SLR methodology, using information from a thermistor and an oximeter, widely outperformed a 6-channel RP. These results suggest that the spectral information from AF can be used to help in pediatric OSAHS detection.

REFERENCES

[1] S. P. Patil, H. Schneider, A. R. Schwartz et al., "Adult Obstructive Apnea," *Chest*, vol. 132(1), pp. 325-337, 2007.
[2] M. S. Schechter, "Technical report: diagnosis and management of childhood obstructive sleep apnea syndrome," *Pediatrics*, vol. 109 (4) pp. e69-e69, 2002.
[3] C. Guilleminault, J. H. Lee, A. Chan, "Pediatric obstructive sleep apnea syndrome." *Arch Pediatr Adolesc Med*, vol. 159, pp. 775-85, 2005.
[4] J. A. Bennet and W. J. M. Kinnear, "Sleep on the cheap: the role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome," *Thorax* vol. 54, pp. 958-959, 1999.
[5] M. L. Álvarez, J. Terán et al., "Reliability of respiratory polygraphy for the diagnosis of sleep apnea-hypopnea syndrome in children," *Arch Bronconeumol*, vol. 44, pp. 318-23, 2008.
[6] J. V. Marcos, R Hornero, D. Álvarez et al., "Automated detection of obstructive sleep apnea syndrome from oxygen saturation recordings using linear Discriminant analysis," *Med Eng Phys*, vol. 59(1), pp. 141-49, 2010.
[7] E. Gil, R. Bailón, J. M. Vergara, et al., "PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children," *IEEE Trans Biomed Eng*, vol. 57(5), pp. 1079-1088, 2010.
[8] D. Alvarez, R. Hornero, J. V. Marcos, et al., "Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis," *Int J Neural Syst*, vol. 23(5), pp. 1-18, 2013.
[9] A. Garde, W. Karlen, P. Dehkordi, et al., "Oxygen saturation in children with and without obstructive sleep apnea using the phone-oximeter," in *Proc. 35th Annual International Conference of the IEEE Engineering and Biology Society*, pp.2531-34, 2013.
[10] 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.
[11] G. C. Gutiérrez-Tobal, D. Álvarez, J. V. Marcos et al., "Pattern recognition in airflow recordings to assist in the sleep apnoea–hypopnoea syndrome diagnosis," *Med Biol Eng Comput*, vol. 51, pp. 1367-80, 2013.
[12] G. C. Gutiérrez-Tobal, R. Hornero, D. Álvarez, et al., "Apnea-hypopnea index estimation from spectral analysis of airflow recordings," in *Proc. 34th Annual International Conference of the IEEE Engineering and Biology Society*, pp.3444-47, 2012.
[13] 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 Trans on Audio and Electroacoustics*, vol. AU-15, pp. 70-73, 1967.
[14] D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression*. John Wiley and Sons, New York, 2000.
[15] I. H. Witten, E. Frank, M. A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann/Elsevier, Burlington, MA, 2011.
[16] E. Tabachnik, N. Muller, B. Toye et al., "Measurement of ventilation in children using the respiratory inductive plethysmograph," *J Pediatr*, vol. 99(6), pp. 895-99, 1981.
[17] R. E. Shouldice, L. M. O'Brien, C. O'Brien et al., "Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features," *Sleep*, vol. 27(4), pp. 784-92, 2004.