

Bispectral Analysis to Enhance Oximetry as a Simplified Alternative for Pediatric Sleep Apnea Diagnosis

Gonzalo C. Gutiérrez-Tobal, *Member, IEEE*, Leila Kheirandish-Gozal, Fernando Vaquerizo-Villar, Daniel Álvarez, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, and Roberto Hornero, *Senior Member, IEEE*

Abstract— This study aims at assessing the bispectral analysis of blood oxygen saturation (SpO_2) from nocturnal oximetry to help in pediatric sleep apnea-hypopnea syndrome (SAHS) diagnosis. Recent studies have found excessive redundancy in the SAHS-related information usually extracted from SpO_2 , while proposing only two features as a reduced set to be used. On the other hand, it has been suggested that SpO_2 bispectral analysis is able to provide complementary information to common anthropometric, spectral, and clinical variables. We address these novel findings to assess whether bispectrum provides new non-redundant information to help in SAHS diagnosis. Thus, we use 981 pediatric SpO_2 recordings to extract both the reduced set of features recently proposed as well as 9 bispectral features. Then, a feature selection method based on the fast correlation-based filter and bootstrapping is used to assess redundancy among all the features. Finally, the non-redundant ones are used to train a Bayesian multi-layer perceptron neural network (BY-MLP) that estimate the apnea-hypopnea index (AHI), which is the diagnostic reference variable. Bispectral phase entropy was found complementary to the two previously recommended features and a BY-MLP model trained with the three of them reached high agreement with actual AHI (intra-class correlation coefficient = 0.889). Estimated AHI also showed high diagnostic ability, reaching 82.1%, 81.9%, and 90.3% accuracies and 0.814, 0.880, and 0.922 area under the receiver-operating characteristics curve for three common AHI thresholds: 1 e/h, 5 e/h, and 10 e/h, respectively. These results suggest that the information extracted from the bispectrum of SpO_2 can improve the diagnostic performance of the oximetry test.

I. INTRODUCTION

Pediatric sleep apnea-hypopnea syndrome (SAHS) has become a main research subject in both medical and technical fields [1]-[4]. On the one hand, the recurrence of respiratory pauses (apneas) and airflow reductions (hypopneas) during the night triggers a sequence of undesirable physiological phenomena that include inadequate gas exchange, oxygen desaturations, arousals, and sleep fragmentation [2]. These lead the affected children to restless sleep, which often results in cardiovascular and metabolic malfunctioning as well as

cognitive and behavioral problems that diminish their health and quality of life [1]. On the other hand, nocturnal polysomnography (PSG), which is the standard diagnostic test, is technically complex due to the number of biomedical signals to be recorded, time-consuming because of the need for inspecting all of them, and it is also costly for healthcare systems since it requires overnight in-hospital supervision [4]. In addition, it is an uncomfortable test to undertake for children [4].

The negative impact of SAHS in children's life accentuates the importance of a rapid diagnosis. However, its high prevalence [1], together with the above-mentioned PSG limitations, lead most of affected children to remain undiagnosed [5], as well as highlights the need for a simplified diagnostic test. Nocturnal pulse-oximetry (NPO) has been often used for this purpose [3], [4], [6]-[8].

NPO is able to record the oxygen saturation signal (SpO_2), one of the 32 signals involved in PSG, with the only use of a sensor placed on a finger [9]. Apneic events usually cause drops in the oxygen of the hemoglobin present in the blood, i. e., oxygen desaturations [9]. Hence, SpO_2 carries useful information about pediatric SAHS while changing the diagnostic approach from 32 channels to a single one. In addition, an automatic methodology has been commonly adopted as a further step into SAHS diagnosis simplification [3], [4], [6]-[8].

A number of studies have shown the usefulness of different automatic approaches, such as spectral and non-linear analyses, when combining them with machine-learning methods to predict the presence of pediatric SAHS as well as its severity [3], [4], [6]-[8]. However, a recent study involving 4,191 pediatric subjects has shown significant redundancy between the information extracted by spectral and non-linear methods and the 3% oxygen desaturation index (ODI_3) [4], a commonly used clinical parameter. By contrast, another study has pointed to bispectral analysis as complementary to ODI_3 and other anthropometric variables [7], suggesting the

*This work was supported 'Ministerio de Economía y Competitividad' and 'European Regional Development Fund' under projects DPI2017-84280-R, TEC2014-53196-R, and RTC-2015-3446-1, by 'Consejería de Educación de la Junta de Castilla y León and FEDER' under project VA037U16, by 'European Commission' and 'European Regional Development Fund' under project 'POCTEP 0378_AD_EEGWA_2_P', and the project 66/2016 de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR).

F. Vaquerizo-Villar was in receipt of a 'Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)' grant from the Ministerio de Educación, Cultura y Deporte. V. Barroso-García was in a receipt of a 'Ayuda para financiar la contratación predoctoral de personal investigador' grant from the Consejería de Educación de la Junta de Castilla

y León and the European Social Fund. D. Álvarez was in receipt of a Juan de la Cierva grant from MINECO. L. Kheirandish-Gozal was supported by National Institutes of Health grant HL130984.

G. C. Gutiérrez-Tobal, F. Vaquerizo-Villar, V. Barroso-García and R. Hornero, are with the Biomedical Engineering Group, Universidad de Valladolid, Spain (e-mail: gonzalo.gutierrez@gib.tel.uva.es).

D. Álvarez, F. del Campo, and A. Crespo are with the Hospital Universitario Río Hortega of Valladolid, Spain (e-mail: fsas@telefonica.net).

L. Kheirandish-Gozal and D. Gozal are with the Section of Sleep Medicine, Dept. of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, Chicago, IL, 60637, USA (e-mail: dgozal@uchicago.edu).

information obtained from the phase of the signal as the source of this new evidence.

According to the above-mentioned considerations, we hypothesize that the information obtained from the bispectrum of the SpO₂ signal is able to complement that provided by the optimum feature subset previously reported [4], i.e., ODI_3 and the third-order statistical moment obtained from the 0.020-0.044 Hz spectral band ($Mf\beta_{Bol}$) [4]. In this regard, the objective of this preliminary study is twofold. First, to evaluate whether 9 bispectral SpO₂-related features show redundancy with ODI_3 and $Mf\beta_{Bol}$. Then, to assess the diagnostic usefulness of the bispectrum features together with the previously reported optimum feature set. A Bayesian neural network approach with ability to automatically estimate pediatric SAHS severity has been chosen for this purpose in view of its previous usefulness in SAHS context [10], [11].

II. SUBJECTS AND SIGNALS

Nine hundred and eighty one SpO₂ recordings were involved in this study. All of them were acquired from children referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children's Hospital (Chicago, IL, USA) due to suspicion of SAHS. The Ethical Committee approved the protocol and all the legal guardians of the pediatric subjects signed and informed consent. Children underwent overnight PSG, which was used as the diagnostic reference. A clinical specialist inspected all PSGs to determine the apnea-hypopnea index (number of apneas and hypopneas per hour of sleep, AHI), the parameter used to determine SAHS and its severity. Apneas and hypopneas were scored following the rules of the American Academy of Sleep Medicine (AASM) [9]. One hundred and seventy five pediatric children showed no SAHS (AHI<1 events per hour). By contrast, 401 had an AHI in the range [1,5) e/h, 176 in the range [5,10) e/h, and 229 showed an AHI ≥ 10 e/h. The whole group was randomly divided into a training set (60%) and a test set (40%). Table I displays clinical and demographic data from the subjects under study for the whole sample as well as split into these two sets. No statistically significant differences (p -value>0.01) were found between training and test samples in either age or body mass index (BMI).

The SpO₂ recording of each child was acquired during its corresponding PSG at sampling rates of 25, 200, and 500 Hz. Hence, a pre-processing stage was included to resample all of them to 25 Hz., as recommended by the AASM [9]. In addition, artifacts due to children movements were removed following the methodology suggested in previous studies [4], [12].

III. METHODOLOGY

Eleven features were obtained from each SpO₂ recording. Two of them (ODI_3 and $Mf\beta_{Bol}$) were reported as the only non-redundant features in a previous study involving 23 SpO₂ features from 4,191 pediatric recordings [4]. This conclusion was reached using the same training set than in the present study. Additionally, 9 features were also extracted from the bispectrum of the SpO₂ recordings. After feature extraction, a feature selection stage was implemented to assess whether there was redundancy among the 11 features obtained. Finally, those features showing no redundancy were used to train and test a Bayesian multi-layer perceptron neural network

TABLE I. DEMOGRAPHIC AND CLINICAL DATA OF THE SUBJECTS UNDER STUDY.

Data	All	Training	Test
Subjects (n)	981	589	392
Males (%)	61.4	58.9	65.1
Age (years)	6.1 \pm 3.4	6.1 \pm 3.4	6.1 \pm 3.5
BMI (kg/m ²)	19.7 \pm 7.3	19.7 \pm 7.5	19.6 \pm 7.0

BMI: Body Mass Index

(BY-MLP) with ability to automatically estimate the AHI of each subject.

A. Feature extraction

- ODI_3 is a widely used clinical parameter commonly involved in SAHS context. ODI_3 is computed as the number of drops from the baseline ($\geq 3\%$) of the SpO₂ time series, divided by the number of hours of the recording [12]. ODI_3 tends to increase with the number of apneic events [9].
- $Mf\beta_{Bol}$ is the third order statistical moment (or skewness) obtained in the frequency domain from the spectral band of interest 0.020-0.044 Hz. It measures the asymmetry of the values of the power spectral density in such band, which has been reported to be different in SAHS affected children [4].
- *Bispectral features*. In contrast to the conventional spectral analysis, as that used to compute $Mf\beta_{Bol}$, bispectrum gathers information not only from the amplitude of a time series but phase too [13]. It is the 2-D Fourier's transform of the third order cumulant and, therefore, it is often represented as a 2-D image or a matrix [13]. Since recurrent drops in SpO₂ changes both the amplitude and phase of the signal, which otherwise shows low variation in the range 96%-100%, up to 9 bispectrum features has been proposed to characterize SAHS information [7]. Features related to amplitude are: mean amplitude of the bispectrum ($MB1$); bispectral entropy ($BE1$) and squared bispectral entropy ($BE2$); and the sum of logarithmic amplitudes ($H1$), sum of the logarithmic amplitudes of the elements of the diagonal of bispectrum ($H2$), and first-order spectral moment of the logarithmic amplitudes of the elements in the diagonal of bispectrum ($H3$). On the other hand, features related to phase are phase entropy (PE) and the mean and the variance of the bispectral invariant ($P(a)$), $meanPa$ and $varPa$, which is the phase of the integration function of the bispectrum for each pair of frequencies associated through a line of slope a [13].

B. Feature selection

The fast correlation-based filter (FCBF) was used to evaluate the redundancy of the bispectral features with ODI_3 and $Mf\beta_{Bol}$. FCBF computes symmetrical uncertainty between each feature, F_i , and the AHI, which is taken as a reference variable ($SU_{F_i,AHI}$) [14]. SU is used to rank higher those features with values closer to 1. Redundancy is established by computing SU between each pair of features (SU_{F_i,F_j}), where F_i is ranked higher than F_j . Then, if $SU_{F_i,F_j} \geq SU_{F_i,AHI}$, F_j is considered redundant and discarded [14]. In order to conduct a comprehensive redundancy analysis and obtain a stable optimum feature subset [15], a bootstrap methodology was used along with the FCBF. Hence, following the methodology

conducted in [4], 1000 new bootstrap sets were formed by resampling with replacement from the training set [16]. Those features selected more often than discarded, i.e., more than 500 times, were chosen as non-redundant [4].

C. Bayesian multi-layer perceptron

A BY-MLP neural network was used to estimate the AHI of the subjects under study. An array formed with the values of the selected features was used as pattern to characterize each subject and train the model. The BY-MLP was arranged in three layers of neurons (input, hidden, and output layer). This architecture configuration is known to be able to provide universal approximations [17]. The weights connecting neurons from different layers were estimated using the Bayesian approach. It showed higher performance than the maximum likelihood optimization approach in previous studies involving SAHS [10]. The number of neurons in the input layer will equal the number of selected features, whereas the output layer will be formed by a single neuron since the model is intended for regression [17]. In order to control the effective network complexity and deal with overfitting [17], the number of neurons in the hidden layer (N_H) was set to 10% of the samples in the training set ($N_H=60$).

D. Statistical analysis

A Bland-Altman plot and the intra-class correlation coefficient (ICC) were used to measure the agreement between the AHI estimated by the BY-MLP methodology and the actual AHI obtained from the standard diagnostic test. The diagnostic performance of the estimated AHI was assessed in terms of sensitivity (Se), specificity (Sp), accuracy (Acc) and area under the receiver-operating characteristics curve (AROC) in three commonly used AHI thresholds (1 e/h, 5 e/h and 10 e/h) [4].

IV. RESULTS

A. Complementarity of the extracted features

Fig. 1 shows a histogram with the number of times each extracted feature was selected by the FCBF after the bootstrap procedure applied to the training set. In accordance with previous studies [4], the clinical parameter ODI_3 was selected all the times and $Mf3_{Bol}$ was also selected more than half of the times. Notice, however, that an additional feature from bispectrum (PE) was also selected more than half of the times. We also verified that PE and $Mf3_{Bol}$ were jointly selected 407 out of the 1000 times. Therefore, ODI_3 , $Mf3_{Bol}$, and PE were chosen as the optimum feature subset.

B. Agreement between estimated and actual AHI

A BY-MLP model was obtained using the three selected features from the subjects of the training set. Fig. 2 shows a Bland-Altman plot with the AHI from the subjects in the test set estimated with this model and the corresponding actual AHI. A slight overestimation is observed in our proposal (mean = 0.725 e/h) with overall low variance in the comparison between estimated and actual AHI (standard deviation = ± 6.21). This results in a 95% confidence interval ranging [-11.45, 12.90] e/h, with higher AHI values (from 10 e/h onwards) being responsible for most of the dispersion in the data. High agreement between estimated and actual AHI is also showed by the ICC value, which reaches 0.889.

C. Diagnostic performance of the estimated AHI

Table II shows the diagnostic performance of our estimation in terms of Se, Sp, Acc, and AROC in three clinically-used AHI thresholds: 1 e/h, 5 e/h, and 10 e/h. High overall diagnostic ability is reached (all Acc values higher than 80.0% and AROC values higher than 0.800), with increasing overall statistics as the AHI threshold is higher.

V. DISCUSSION AND CONCLUSIONS

In this preliminary study, we assessed the usefulness of the

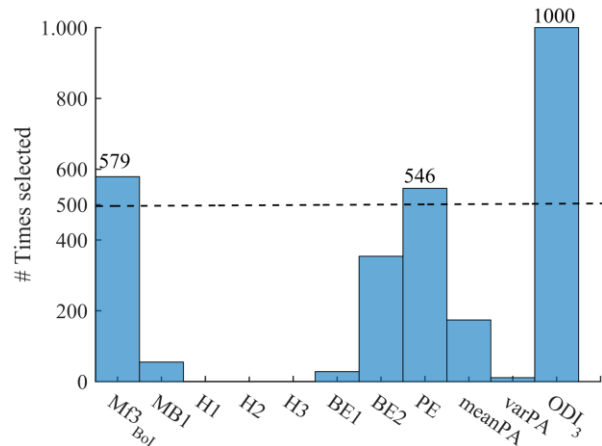


Figure 1. Histogram of the number of times each features was selected (bootstrap conducted in the training set).

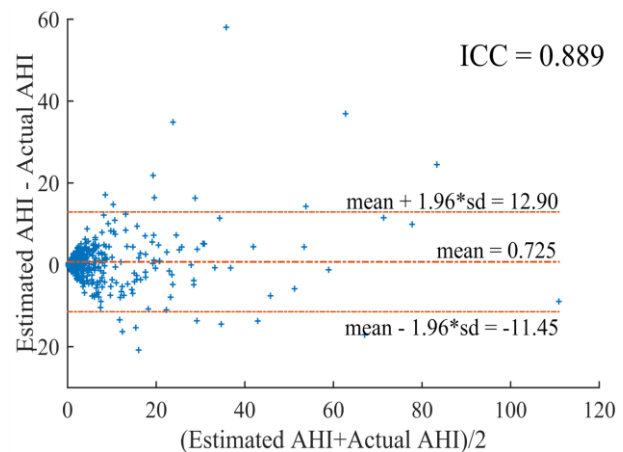


Figure 2. Bland-Altman plot with the estimated and actual AHI in the test set (sd stands for standard deviation).

TABLE II. DIAGNOSTIC ABILITY OF THE ESTIMATED AHI FOR THE AHI THRESHOLDS 1 E/H, 5 E/H AND 10 E/H (TEST SET)

AHI threshold	Se (%)	Sp (%)	Acc (%)	AROC
1 e/h	93.0	37.7	82.1	0.814
5 e/h	78.1	84.2	81.9	0.880
10 e/h	75.9	94.2	90.3	0.922

bispectral analysis of the SpO₂ signal in the simplification of the diagnosis of pediatric SAHS. We found that the entropy of the bispectrum phase showed complementarity with ODI_3 and $Mf\beta_{Bot}$, which had been recently established as relevant and non-redundant features for pediatric SAHS in a multicenter study involving 4,191 subjects [4]. A BY-MLP neural network was trained with these three features to automatically estimate the AHI of the subjects under study. It was subsequently tested, reaching high agreement with actual AHI (ICC=0.889) and high diagnostic ability for 3 commonly-used AHI thresholds (1 e/h, 5 e/h, and 10 e/h). Particularly high were the figures for 5 e/h and 10 e/h, reaching 0.880 and 0.922 AROC.

Only one study has been found focused on estimating AHI from SpO₂ features from pediatric recordings. Hornero *et al.* involved 4,191 subjects and obtained a multilayer perceptron model built with ODI_3 and $Mf\beta_{Bot}$ [4]. They reported an agreement with actual AHI of 0.785 ICC. Additionally, they reached 75.2%, 81.7%, and 90.2% Acc as well as 0.788, 0.854, and 0.913 AROC for the AHI thresholds 1 e/h, 5 e/h, and 10 e/h, respectively. Our new proposal outperformed all these diagnostic statistics, with higher gain for the threshold of 1 e/h. However, they used 3,602 subjects as test set in contrast to 392 in this study.

Vaquerizo-Villar *et al.* used 2 SpO₂ bispectral features along with 7 anthropometric, spectral, and clinical features to train a multi-layer perceptron neural network with ability to classify pediatric subjects in 3 classes (AHI<5 e/h, AHI in the range [5,10), and AHI≥10 e/h) [7]. Using a database with 75 subjects in the test set, they reported 81.3% and 85.3% Acc for AHI thresholds of 5 e/h and 10 e/h, respectively. Finally, a number of studies took a binary approach by training models with ability to discriminate pediatric SAHS in a single AHI threshold, usually 5 e/h [3], [6], [8], [18]. Accuracies reported in these works range 76.6%-84.9% using relatively small databases (from 21 to 146 pediatric subjects) [3], [6], [8], [18]. Our AHI estimation proposal, which can be used and evaluated in several thresholds at the same time, ranked high among these works only focused on the AHI threshold equal to 5 e/h [3], [6], [8], [18].

In spite of the promising results showed in this study, several limitations need to be addressed. First, further evaluation would be required in order to fairly compare our results with those reported in the study by Hornero *et al.* In this regard, the evaluation of our new proposal in the 4,191 subjects is a future goal. In addition, other features different from bispectral analysis may also gather complementary information to simplify SAHS diagnosis through SpO₂ signal. Hence, other analytical approaches, such as time-frequency analysis, could help to gain insight into the maximal diagnostic capability of this signal. Finally, the number of hidden neurons in the BY-MLP network was arbitrary chosen to control the effective complexity of the network. Therefore, additional tests would be required to find the optimum number of hidden neurons for our problem.

Summarizing, we found that the phase of the SpO₂

bispectrum provides complementary information to that recently established as relevant and non-redundant for pediatric SAHS diagnosis simplification. It also showed its diagnostic usefulness by taking part of a BY-MLP estimation model that reached high agreement with actual AHI as well as high diagnostic ability comparing with the state-of-the-art studies. These results suggest that the entropy of the bispectrum phase from SpO₂ might be used to enhance the models described in the literature to help in pediatric SAHS simplification.

REFERENCES

- [1] C. L. Marcus, et al., "Diagnosis and management of childhood obstructive sleep apnea syndrome," *Pediatrics*, vol. 130, pp.e714-e755, 2012.
- [2] S. J. Hunter, et al., "Effect of sleep-disordered breathing severity on cognitive performance measures in a large community cohort of young school-aged children," *Am. J. Respir. Crit. Care Med.*, vol. 194, pp. 739-747, 2016.
- [3] A. Garde, et al., "Development of a Screening Tool for Sleep Disordered Breathing in Children Using the Phone Oximeter™," *PLoS one*, vol. 9, pp. e112959, 2014.
- [4] R. Hornero, et al., "Nocturnal Oximetry-based Evaluation of Habitually Snoring Children," *Am. J. Respir. Crit. Care Med.*, vol. 196, pp. 1591-1598, 2017.
- [5] R. A. Weatherly, et al., "Identification and evaluation of obstructive sleep apnea prior to adenotonsillectomy in children: a survey of practice patterns," *Sleep medicine*, vol. 4, 297-307, 2003.
- [6] D. Álvarez, et al., "Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry: An Alternative to Respiratory Polygraphy in Unattended Settings," *J Clin Sleep Med*, vol. 13, pp. 693-702, 2017.
- [7] F. Vaquerizo-Villar, et al., "Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings," *Comp. Meth. Prog. Biomed.*, vol. 156, pp. 141-149, 2018.
- [8] E. Gil, 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, pp. 1079-1088, 2010.
- [9] R. B. Berry, et al., "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events," *J Clin Sleep Med*, vol. 8, pp. 597-619, 2012.
- [10] J. V. Marcos, et al., "The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome," *Phys. Meas.*, vol. 31, pp. 375-94, 2010.
- [11] G. C. Gutiérrez-Tobal, et al., "A Bayesian neural network approach to compare the spectral information from nasal pressure and thermistor airflow in the automatic sleep apnea severity estimation," *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3741-3744, 2017.
- [12] U. J. Magalang et al., "Prediction of the apnea-hypopnea index from overnight pulse oximetry," *Chest*, vol. 124(5), pp. 1694-1701, 2003.
- [13] K.C. Chua, et al., "Application of higher order statistics/spectra in biomedical signals-A review", *Med. Eng. Phys.*, vol. 32, pp. 679-89, 2010.
- [14] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205-1224, 2004.
- [15] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J Mach Learn Res*, vol. 3, pp. 1157-1182, 2003.
- [16] I. H. Witten, et al., *Data Mining Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann/Elsevier, 2011.
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.
- [18] L. Chang, et al., "Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea," *Int. J. Pediatr. Otorhino-laryngol.*, vol. 77, pp. 365-371, 2013.