Feasibility of Unsupervised Sleep Apnea Screening Using Pulse Rate Oscillations and SpO_2 with a Wrist-Worn Device

Diego Cajal[1,2], Rodrigo Lozano-Puñet[1,2], Ana Victoria Gil[3,4], José Pablo Cubero[3,4], Jose María Marín[3,4], Eduardo Gil[1,2], Jesús Lázaro[1,2], Raquel Bailón[1,2]

[1] Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Spain
[2] CIBER - Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain
[3] Hospital Universitario Miguel Servet, Spain
[4] Instituto de Investigación Sanitaria de Aragón, Spain

Abstract

In a previous work, a method was presented to detect sleep apnea using signals commonly found in wearable devices: peripheral oxygen saturation (SpO_2) and pulse photoplethysmography (PPG). However, these signals were obtained from a conventional fingertip pulse oximeter. This paper describes a pilot study applying those methods with signals acquired by a wrist-worn wearable device in a non-supervised home environment ($n = 12$). A classifier was applied to differentiate normal from abnormal breathing segments. Later, the Cyclic Variation of Heart Rate Index (CVHRI) was calculated within the abnormal breathing segments. The classifier achieved an accuracy of 65.7% on the wearable data and CVHRI maintained a strong correlation with the AHI ($r = 0.85$, $p < 0.001$), suggesting its potential for patient stratification remains viable.

1. Introduction

Sleep apnea-hypopnea syndrome (SAHS) is a sleep disorder characterized by recurrent episodes of either breathing interruption or significant reduction in airflow, which can lead to serious health outcomes. These effects have considerable socioeconomic impacts, but they can be mitigated with appropriate treatment. However, SAHS is a common but often underdiagnosed condition, partly due to the limitations of the gold-standard polysomnography [1]. Failure to diagnose SAHS is associated with a higher risk of hypertension, cardiovascular disease, stroke, daytime sleepiness, motor vehicle accidents, and reduced quality of life [2].

Wearable devices offer a promising avenue for accessible, in-home screening. Previous research demonstrated a method using SpO_2 and PPG-derived pulse-to-pulse intervals (PPI) from non-wearable sensors to classify breathing segments into *normal* and *abnormal breathing* segments [3]. The classification achieved satisfactory accuracy, and the PPI from *abnormal breathing* segments was utilized to calculate the CVHRI, as introduced in the same study. This index measures the frequency of the bradycardia-tachycardia pattern that usually appears during repetitive apnea events, demonstrating a strong correlation with the Apnea-Hypopnea Index (AHI).

This work presents a pilot study evaluating the translation of that method to a real-world scenario using data acquired from a wrist-worn wearable device operated by patients and controls unsupervised at home. To maintain the consistency of covariates, the original methodology has been preserved, including methods for event labeling and their conversion to segment labeling. A classifier utilizing Hjorth parameters [4] of PPI and SpO_2 as inputs yielded the best results and is consequently tested here. Furthermore, the model has not been retrained with the new signals but used as it was originally trained on the previous non-wearable database. In this case, the PPI and SpO_2 signals are obtained from the wearable device. CVHRI is also calculated in the same way, using the PPI of the wearable device within *abnormal breathing* segments.

2. Methods

2.1. Experiment

A group of 12 participants (age 49$\pm$16 years, including 7 males) from a larger ongoing study (approved by Aragon's ethics committee, CEICA, PI23/336) was analyzed. This subset includes 8 subjects suspected of having SAHS and 4 control cases. Participants were instructed to use the recording devices over one night in their homes, activating them at bedtime and deactivating them upon waking. The recordings were performed using a commercially available polygraph: the ApneaLink Air device (ResMed Inc., USA), which provided SpO_2 (denoted as

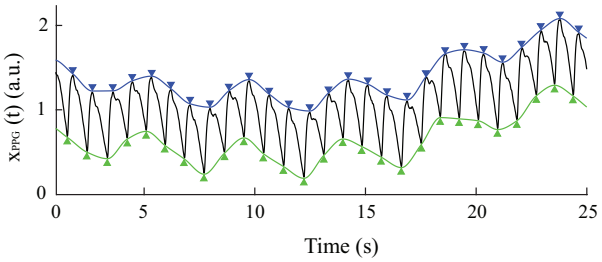

Figure 1. Upper (blue) and lower (green) envelopes of the PPG (black). Blue triangles represent maximum pulse amplitudes, while green triangles represent pulse basal points.

$x_{\mathrm{SpO_2}}[n]$ in this paper) with a sampling rate of 1 Hz and nasal pressure at 100 Hz signals as a reference; and a Maxim MAXREFDES103 device (Analog Devices Inc., USA) as a wrist-worn wearable, which captured PPG signals at 256 Hz with red, green, and infrared channels (denoted $x^r_{\mathrm{PPG}}[n]$, $x^g_{\mathrm{PPG}}[n]$ and $x^{ir}_{\mathrm{PPG}}[n]$, respectively, in this paper).

Reference signals from the Apnealink device were used exclusively to label apneic and hypopneic events. These events were annotated according to AASM scoring guidelines [5], implementing an automated algorithm detailed in [3]. Subsequently, event-based annotations were aggregated into segment-based annotations to act as classifier labels [3]. These annotations were then compared to the classifier output using the wearable's signals as input. In addition, the reference was used to calculate the AHI, which was compared to the CVHRI.

2.2. SpO_2 estimation

PPI and SpO_2 were estimated from the wearable PPG signal and denoted $d_{\mathrm{PPI}}[n]$ and $d_{\mathrm{SpO_2}}[n]$, respectively, in this paper. $d_{\mathrm{PPI}}[n]$ was computed from $x^g_{\mathrm{PPG}}[n]$ by identifying the maximum upslope keypoints, as outlined in [6]. To calculate $d_{\mathrm{SpO_2}}[n]$, the R ratio, $d_{\mathrm{R}}[n]$, has been estimated. This ratio is derived from the AC/DC ratios of $x^r_{\mathrm{PPG}}[n]$ and $x^{ir}_{\mathrm{PPG}}[n]$:

$$d_{\mathrm{R}}[n] = \frac{\frac{d^{r}_{\mathrm{AC}}[n]}{d^{r}_{\mathrm{DC}}[n]}}{\frac{d^{ir}_{\mathrm{AC}}[n]}{d^{ir}_{\mathrm{DC}}[n]} + \epsilon}, \tag{1}$$

where a small value $\epsilon = 0.01$ has been added to the denominator. This is needed as the signal amplitude may temporarily decrease to near-zero values due to movements that lead to suboptimal placement of the sensor.

The AC component of each channel is calculated as the difference between the upper and lower envelopes of $x^r_{\mathrm{PPG}}[n]$ and $x^{ir}_{\mathrm{PPG}}[n]$ (see Figure 1). These envelopes were computed by interpolation of maximum pulse amplitudes and pulse basal points defined as in [6], respectively, at 256 Hz. The DC component is the lower envelope.

The errors are then removed using a fixed threshold applied to the squared derivative of $d_{\mathrm{R}}[n]$. This threshold is empirically set at 10^{-8}, effectively eliminating outliers caused primarily by envelope estimation errors. A second threshold is applied to $d_{\mathrm{R}}[n]$ to remove values that exceed 4. Subsequently, the signal is low-pass filtered at 0.1 Hz to eliminate rapid variations caused by noise.

According to the wearable documentation, oxygen saturation should be calculated using the formula:

$$d_{\mathrm{SpO_2}}[n] = -16.6(d_{\mathrm{R}}[n])^2 + 8.3(d_{\mathrm{R}}[n]) + 100. \tag{2}$$

However, this formula did not yield satisfactory results with our data. Alternatively, a subject-specific calibration was carried out by linearly transforming $d_{\mathrm{R}}[n]$ to have the same mean and standard deviation as the reference $x_{\mathrm{SpO_2}}[n]$. Finally, the signal was resampled at 25 Hz to comply with the recommendations of the AASM [5] and rounded to the nearest integer (see Figure 2).

2.3. Segment classification and CVHRI

$d_{\mathrm{PPI}}[n]$ and $d_{\mathrm{SpO_2}}[n]$ derived from the wearable device were utilized to identify segments corresponding to *normal* and *abnormal breathing*. First, the signals were divided into 180-second segments with a 150-second overlap. From each segment, Hjorth parameters (Activity, Mobility, and Complexity) were extracted from both $d_{\mathrm{PPI}}[n]$ and $d_{\mathrm{SpO_2}}[n]$. This process yields six parameters per segment, which serve as input features for a classifier. Notably, this classifier was previously trained using signals originating from non-wearable devices, as detailed in [3]. Given that, in this instance, $d_{\mathrm{SpO_2}}[n]$ was estimated from the wearable PPG, the classifier was evaluated utilizing both this estimation and, separately, $x_{\mathrm{SpO_2}}[n]$ from the reference device. This comparative approach allows for the measurement of how this particular estimation technique affects the classifier's performance.

Subsequently, the CVHRI is computed from $d_{\mathrm{PPI}}[n]$ within those segments classified as exhibiting *abnormal breathing*. It is within these particular segments that the characteristic tachycardia-bradycardia pattern is expected to manifest. The CVHRI is defined as the sum of the frequencies of the spectral peaks, divided by the total number of segments analyzed. This methodology results in a single, patient-specific parameter, analogous to the AHI.

3. Results

The annotation using reference signals resulted in a total of 6,143 *normal breathing* segments and 1,109 *abnormal breathing* segments. The mean AHI was 5.95, with a standard deviation of 6.67. 7 subjects had AHI < 5; 3 subjects had $5 \leq$ AHI < 15; and 2 subjects had AHI ≥ 15.

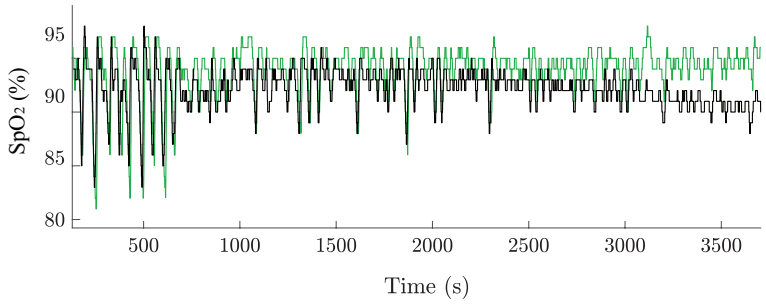

Figure 2. SpO_2 estimation from Maxim. SpO_2 from reference in green, estimate from Maxim in black.

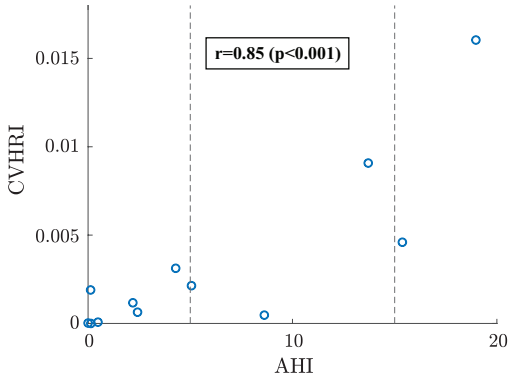

Figure 3. AHI vs CVHRI. $d_{\text{SpO}_2}[n]$ estimated from the wearable. Vertical lines separate AHI $<$ 5, 5$\leq$AHI $<$ 15 and AHI $\geq$ 15 groups.

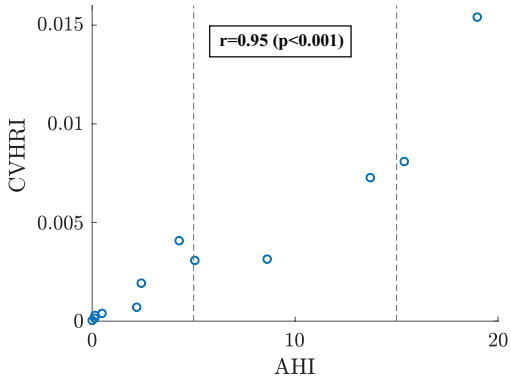

Figure 4. AHI vs CVHRI. $x_{\text{SpO}_2}[n]$ from the reference. Vertical lines separate AHI $<$ 5, 5$\leq$AHI $<$ 15 and AHI $\geq$ 15 groups.

Following random class balancing [3], the classification results for the segments are consolidated in Table 1. Columns Acc, P_n, R_n, P_{ab} and R_{ab} represent the classifier accuracy, precision for the *normal breathing* class, recall for the *normal breathing* class, precision for the *abnormal breathing* class, and recall for the *abnormal breathing* class, respectively. CVHRI correlates with the reference AHI, with a Pearson correlation coefficient of $r = 0.85$ ($p < 0.001$). Figure 3 presents a scatter plot of AHI vs CVHRI.

To assess the impact of estimating $d_{\text{SpO}_2}[n]$, a new CVHRI is calculated from the outcomes of the segment classification utilizing both the wearable $d_{\text{PPI}}[n]$ and the reference $x_{\text{SpO}_2}[n]$. Obtained correlation with AHI was $r = 0.95$, $p < 0.001$ (see Figure 4).

Table 1. Segment classification results (%).

$\mathbf{SpO_2}$	$\boldsymbol{Acc}$	$\boldsymbol{P_n}$	$\boldsymbol{R_n}$	$\boldsymbol{P_{ab}}$	$\boldsymbol{R_{ab}}$
Wearable	65.7	87.8	60.9	43.6	78.2
Conventional	77.0	88.1	72.1	65.9	84.7

4. Discussion

The results obtained using the wearable signals show a degradation compared to those calculated with the polysomnography system in [3]. This was expected for several reasons: a noisier PPG signal prone to artifacts; a $d_{\text{SpO}_2}[n]$ estimation derived from this signal; and the use of a pre-trained model designed for other types of signals. The lower robustness of the PPG signal is linked to the in-home non-supervised approach. However, the model could be retrained for this kind of signals when more data is available, and this remains a room for improvement.

Estimation of $d_{\text{SpO}_2}[n]$ from the wearable was not performed using the formula provided by the manufacturer. Instead, a subject-specific calibration was used, which led to good qualitative results. However, future work should include a quantitative evaluation and optimization. The need to calibrate a wearable device for each user is not the optimal situation. However, it is common practice to request prior calibration from users for other types of measurement, such as blood pressure. Nevertheless, the approach taken in this experiment could be improved, potentially enabling better classification results, as evidenced by the results obtained using the reference signal $x_{\text{SpO}_2}[n]$. An

alternative to this approach would be to retrain the classification model using the Hjorth parameters of $d_{\mathrm{R}}[n]$ instead of the estimated $d_{\mathrm{SpO_2}}[n]$. Since the parameters reflect the variance, center frequency, and bandwidth of the signal, the operations performed for estimation (*i.e.*, normalization, multiplication by a constant, and addition of a constant) should lead to proportional values.

The classification model should be re-trained using the new signals as input when more data is available. An improvement in results is expected, given the significant differences between $x_{\mathrm{SpO_2}}[n]$ and $d_{\mathrm{SpO_2}}[n]$.

The number of subjects should also be increased, as the current sample is limited, with only 2 subjects having an AHI ≥ 15. This constrains the analysis that could be performed, excluding the stratification applied in [3]. Moreover, the reduced number of *abnormal breathing* segments means that the classifier's statistics are calculated using 1,109 segments per class, since class balance ensures that each class has the same number of samples as the minority class, in this case, the *abnormal breathing* class. Consequently, the majority of classified segments are excluded from evaluating the classifier's performance. Therefore, more subjects should be recorded, with a particular focus on expanding the subgroup with AHI ≥ 15.

Despite this degradation, the results remain promising. The correlation between CVHRI and AHI remains high, decreasing only slightly from 0.94 with the non-wearable signals [3] to 0.85 with those from the wearable. This likely enables effective stratification for sleep apnea screening. The drop in correlation is primarily attributed to the decline in classifier accuracy, which decreased from 86.3% [3] to 65.7%. However, this pronounced degradation in accuracy does not correspond to an equivalent percentage drop in the AHI-CVHRI correlation, suggesting that CVHRI may be a robust metric against classifier errors.

Nonetheless, the potential for improving $d_{\mathrm{SpO_2}}[n]$ estimation is evident in the results. When $x_{\mathrm{SpO_2}}[n]$ is used instead of the wearable-derived $d_{\mathrm{SpO_2}}[n]$, the correlation returns to 0.95, equivalent to that of [3], and the classifier accuracy improves to 77.0%. While this is not as high as in [3], it surpasses the accuracy achieved when only wearable-derived inputs are used. This outcome is consistent with the findings of [3], where Hjorth parameters of $x_{\mathrm{SpO_2}}[n]$ were shown to be the main predictors.

5. Conclusions

The results suggest that the method studied in [3] can be applied to wearable devices used at the patient's home with no supervision, delivering good performance. The classifier achieved an accuracy of 65.7%, while the correlation between CVHRI and AHI was $r = 0.85$ ($p < 0.001$). The high correlation indicates that stratification remains feasible. However, the sample size should be increased, particularly for the group with AHI ≥ 15, to enable a more comprehensive study. Additionally, efforts should focus on improving the estimation of $d_{\mathrm{SpO_2}}[n]$ from $d_R[n]$ or directly using the latter as an input. Combined with retraining the model to account for the new characteristics of the signals, these steps are expected to enhance the results further.

Acknowledgments

This work is supported by 'Ministerio de Ciencia, Innovación y Universidades, Spain' and 'European Regional Development Fund (FEDER)' under project PID2021-126734OB-C21, by 'CIBER - Consorcio Centro de Investigación Biomédica en Red, Spain' through 'Instituto de Salud Carlos III' co-funded with ERDF funds under project TED2021-131106BI00, and by 'Gobierno de Aragón, Spain' (Reference Group BSICoS T39-23R). Computations were performed by the ICTS NANBIOSIS (HPC Unit at University of Zaragoza).

References

[1] Yeghiazarians Y, Jneid H, Tietjens JR, Redline S, Brown DL, El-Sherif N, Mehra R, Bozkurt B, Ndumele CE, Somers VK. Obstructive sleep apnea and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 7 2021;144(3). ISSN 0009-7322.

[2] Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. American Journal of Respiratory and Critical Care Medicine 5 2002; 165(9):1217–1239. ISSN 1073-449X.

[3] Cajal D, Gil E, Laguna P, Varon C, Testelmans D, Buyse B, Jensen C, Hoare R, Bailon R, Lazaro J. Obstructive sleep apnea screening by joint saturation signal analysis and PPG-derived pulse rate oscillations. IEEE Journal of Biomedical and Health Informatics 2023;ISSN 21682208.

[4] Hjorth B. EEG analysis based on time domain properties. Electroencephalography and Clinical Neurophysiology 1 1970;29(3):306–310.

[5] Berry R, Brooks R, Gamaldo C, Harding SM, Marcus CL, Vaughn BV, Tangredi MM. The AASM Manual for the scoring of sleep and associated events: rules, terminology and technical specifications, version 2. Rules Terminology and Technical Specifications Darien Illinois American Academy of Sleep Medicine 2012;URL `www.aasmnet.org`.

[6] Lázaro J, Gil E, Vergara JM, Laguna P. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. IEEE Journal of Biomedical and Health Informatics 2014;18(1):240–246. ISSN 21682194.

Address for correspondence:

Diego Cajal (dcajal@unizar.es). Universidad de Zaragoza.