

# Automatic Silence Events Detector from Smartphone Audio Signals: A Pilot mHealth System for Sleep Apnea Monitoring at Home

Yolanda Castillo-Escario, Ignasi Ferrer-Lluis, Josep Maria Montserrat, Raimon Jané, *Senior Member IEEE*

**Abstract**— Obstructive sleep apnea (OSA) is a prevalent disease, but most patients remain undiagnosed and untreated. Recently, mHealth tools are being proposed to screen OSA patients at home. In this work, we analyzed full-night audio signals recorded with a smartphone microphone. Our objective was to develop an automatic detector to identify silence events (apneas or hypopneas) and compare its performance to a commercial portable system for OSA diagnosis (ApneaLink™, ResMed). To do that, we acquired signals from three subjects with both systems simultaneously. A sleep specialist marked the events on smartphone and ApneaLink signals. The automatic detector we developed, based on the sample entropy, identified silence events similarly than manual annotation. Compared to ApneaLink, it was very sensitive to apneas (detecting 86.2%) and presented an 83.4% positive predictive value, but it missed about half the hypopnea episodes. This suggests that during some hypopneas the flow reduction is not reflected in sound. Nevertheless, our detector accurately recognizes silence events, which can provide valuable respiratory information related to the disease. These preliminary results show that mHealth devices and simple microphones are promising non-invasive tools for personalized sleep disorders management at home.

## I. INTRODUCTION

Obstructive sleep apnea (OSA) is a common sleep disorder characterized by repetitive breathing interruptions due to upper airways obstruction during sleep. This leads to hypoxia and microarousals, which have been related to an increased cardiovascular and cerebrovascular morbidity and mortality, as well as an increased risk of traffic accidents due to daytime sleepiness [1]. Evidence suggests that OSA prevalence ranges from 9% to 38% in the overall population, being higher in men and reaching values as high as 80% in some elderly groups [2]. Nonetheless, despite the serious health and economic impact of this disease, most patients remain undiagnosed and untreated.

The gold-standard technique for OSA diagnosis is nocturnal polysomnography (PSG). It records many

\*Research funded by fellowships from “la Caixa” Foundation (ID 100010434, fellowship codes LCF/BQ/IN17/11620029, LCF/BQ/ES18/11670019), the EU Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant No. 713673, CERCA Program, Secretaria d’Universitats i Recerca de la Generalitat de Catalunya (GRC 2017 SGR 01770), the Spanish Ministry of Economy and Competitiveness (DPI2015-68820-R MINECO/FEDER) and ISCIII (FIS PI17/01068)

Y. Castillo-Escario, I. Ferrer-Lluis and R. Jané are with the Institute for Bioengineering of Catalonia (IBEC), the Barcelona Institute of Science and Technology (BIST), Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN) and ESII Department, Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Barcelona, Spain (e-mail: {ycastillo, iferrer, rjane} @ibecbarcelona.eu).

J. M. Montserrat is with Hospital Clínic de Barcelona (Sleep Lab, Pneumology Service) and Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES)

physiological signals overnight in hospital to obtain the apnea hypopnea index (AHI): the average number of these episodes per hour. Apneas are defined as flow reductions  $\geq 90\%$  for  $\geq 10$  seconds, while hypopneas are flow reductions  $\geq 30\%$  for  $\geq 10$  seconds, followed by a  $\geq 3\%$  oxygen desaturation or an arousal [3].  $AHI < 5$  means normality,  $5 \leq AHI < 15$  mild OSA,  $15 \leq AHI < 30$  moderate OSA and  $AHI \geq 30$  severe OSA [3].

PSG has important limitations: it is complex, high resource consuming, and AHI is measured in only one night. For this reason, there is an increasing need for alternative equipment, and some approaches have been proposed to diagnose OSA at home using unobtrusive sensors. These devices usually measure fewer signals which can provide high clinical value, such as pulse oximetry or audio. The first has long been studied, since it reflects the effects of interrupted breath as reductions in blood oxygen saturation ( $SpO_2$ ). On the other hand, audio requires only a low-cost sensor (a microphone) and can be used to identify apnea and hypopnea episodes as absence or reduction of sound. Furthermore, acoustic analysis of snoring can elucidate the obstruction mechanisms and help detect OSA patients [1,4].

The relative low cost and high quality of smartphone sensors, and its widespread availability, have led to their multiple mHealth applications, including sleep medicine. Some systems and apps have been proposed to screen or monitor OSA patients at home [5-7]. They integrate different smartphone built-in or external sensors, such as microphones, accelerometers or pulse oximeters, to extract OSA severity indices. However, further studies are needed as they are not yet clinically validated. Moreover, most works use external microphones instead of the smartphone built-in microphone.

In this study, we propose to use a mHealth system consisting of a smartphone built-in microphone and an external pulse oximeter. We had previously characterized the response of this type of microphone in short tests in the lab and at home and found that it has enough resolution to detect low breathing [8-9]. The objective of this work is to develop an automatic detector to identify silence events (apneas or hypopneas) and evaluate its performance. To do that, we compared it to both an expert who manually labeled the smartphone signals and a commercial portable system for sleep apnea diagnosis at home.

## II. MATERIALS AND METHODS

### A. Experiments and Data Acquisition

Full-night recordings from three OSA patients (Table I) were acquired at home. All the experiments were approved by the Hospital Clinic Ethics Committee. Overnight tests were performed simultaneously with our mHealth system and a commercial portable device for sleep apnea diagnosis:

ApneaLink™ Air (ResMed). We used a Samsung Galaxy S5, because it has a high-quality microphone, while being a mid-range smartphone [8-9]. It was fixed on the thorax through an elastic band, in the position suggested by Nakano et al. [10], while ApneaLink was fixed just below, as shown in Fig. 1A.

We used the smartphone built-in microphone to acquire audio signals, and an external microUSB-wired pulse oximeter (Kenek Edge, LionsGate Technologies Inc.) to monitor SpO<sub>2</sub> (Fig. 1B). Sampling frequencies were 48 kHz and 2 Hz, respectively. ApneaLink measured respiratory flow through a nasal cannula, thoracic movement and SpO<sub>2</sub>.

### B. Signal Processing and Analysis

Signals were processed and analyzed offline, using custom-made algorithms in MATLAB® (r2018a, Mathworks Inc.). ApneaLink and smartphone signals were manually synchronized by cutting or zero-padding them. Regions not analyzed by ApneaLink software due to artifacts or low signal magnitude were discarded.

Audio signals were downsampled to 5 kHz, applying an anti-aliasing low-pass filter. A Notch filter was used to remove 50 Hz power-line noise and harmonics. A forward-backward 8<sup>th</sup> order Butterworth band-pass filter (cut-off frequencies: 70-2000 Hz) was applied to keep the bandwidth of respiratory and snoring sounds, while reducing cardiac and high-frequency noise. High-amplitude artifacts were discarded using a variance-based detector. Finally, signals were normalized to the maximum amplitude.

SpO<sub>2</sub> signal artifacts (values lower than 50%, higher than 100%, or sudden changes reflected by more than 2% differences between consecutive samples) were identified and padded with the previous correct value. Desaturations were labeled using an automatic custom-made detector which finds when the signal falls  $\geq 3\%$  below an adaptive baseline.

### C. Automatic Silence Events Detector

We developed an automatic detector to identify silence events (apneas/hypopneas) in smartphone audio signals. Our detector is based on the sample entropy (SampEn) of the audio signal. Sample entropy is a measure of complexity, or regularity, of time-series signals [11]. Given  $m$  (embedding dimension),  $r$  (tolerance), and  $N$  (number of data points), SampEn is the negative natural logarithm of the conditional probability that if 2 sets of consecutive data points of length  $m$  have distances lower than  $r$ , then 2 sets of consecutive data points of length  $m+1$  have also distances lower than  $r$  [11].

The more complex the signal, i.e. the less regular, the higher the entropy and vice versa. In our case, snoring and breathing episodes are more complex than silence regions. Moreover, if the tolerance value  $r$  is fixed, then SampEn also depends on signal amplitude [12]. Consequently, we can use it to extract a kind of signal envelope. The main advantage is that it is less influenced by impulsive artifacts than conventional measures [12].

TABLE I. SAMPLE ANTHROPOMETRIC DESCRIPTION

Subject	Gender	Age	BMI (kg/m <sup>2</sup> )	OSA Severity <sup>a</sup>
1	Male	62	27.8	Moderate
2	Male	61	31.4	Severe
3	Male	56	29.8	Moderate

a. According to ApneaLink™

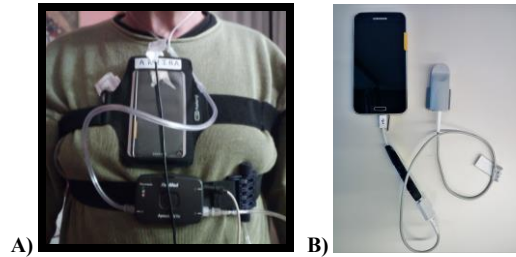


Figure 1. A) Smartphone and ApneaLink position for sleep tests. B) Smartphone with the wired Kenek Edge pulse oximeter (ownership LionsGate Technologies Inc., Vancouver, Canada, [www.lgtmedical.com](http://www.lgtmedical.com), [contact@lgtmedical.com](mailto:contact@lgtmedical.com))

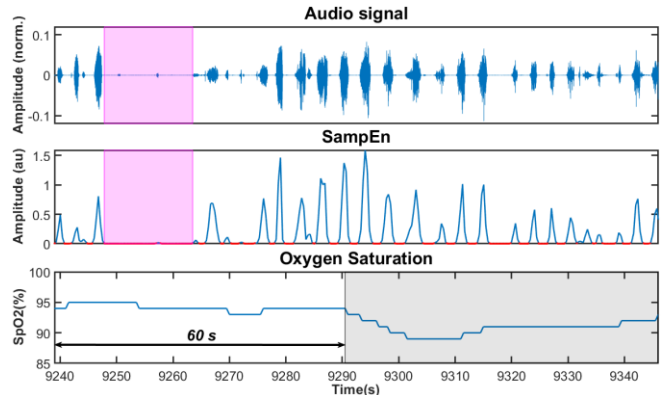


Figure 2. Depiction of the algorithm implementation: a window from 60 s before the desaturation (gray box) is taken, SampEn is computed from the audio signal, values below a threshold are taken (red dots) and regions  $>6s$  associated to desaturations are labeled as “silence events” (magenta boxes).

To reduce the computational cost and false alarms, the SpO<sub>2</sub> signal was used to guide our detector by indicating the regions of interest: those followed by desaturations. Specifically, from each single desaturation event, we selected a window which went from 60 s before the initial time of the desaturation to the end of the desaturation (Fig. 2). To avoid redundancy, overlapping windows were concatenated.

In each segment, the envelope was extracted by computing SampEn. We used 0.75 s windows with 50% overlap, and SampEn parameters were set to  $m=2$  and  $r$  equal to the standard deviation of the segment. Then, points under a certain threshold were selected. Using SampEn strictly equal to 0 might be insufficient in noisy segments, so we set this threshold to 0.01. Points under this value are potential silence regions. They are highlighted with red dots in Fig. 2.

Regions longer than 6 s and shorter than 100 s were selected as candidates to silence events. While apnea/hypopnea definition is a flow reduction  $\geq 10$  s, these events are usually shorter in audio. That is why we set the minimum length to 6 s instead. Events were merged when they were separated less than a distance that would be expected for a breath or snore. Finally, events were directly associated to desaturations. Unpaired events were removed. Those remaining were the final silence events labeled by our detector (Fig. 2).

### D. Manual Labeling & Comparison

An expert in sleep medicine from Hospital Clínic de Barcelona examined smartphone signals (audio and SpO<sub>2</sub>) and was asked to visually identify and label silence events

longer than 6s and followed by a desaturation. ApneaLink channels (flow, thoracic effort and SpO<sub>2</sub>) were separately examined by a sleep technician, who labeled apnea and hypopnea events as in clinical practice. They were used as the gold standard for this study.

Automatic events were compared to both manual smartphone labels and ApneaLink events. Manual labels were also compared to ApneaLink events. The performance of each method was quantified in terms of “event-to-event” agreement. We computed the number of true positives, false positives and false negatives with respect to the reference. Then, sensitivity (Se) and positive predictive value (PPV) were calculated. We also calculated which percentage of each type of event (apnea/hypopnea) we were detecting, i.e. the ratio of events labeled as “apnea” in ApneaLink that we label as “silence event”, and the same for hypopneas.

### III. RESULTS

The automatic detector that we developed identified with precision regions where silence was present in the full-night test signals. Silence events were consistent with those found by visual inspection. Furthermore, they correlated well with ApneaLink events. Fig. 3 shows a representative segment with a high number of apneas and one hypopnea, comparing ApneaLink flow channel and smartphone audio signal. In this example, all the ApneaLink events are identified by our automatic detector.

Table II contains the comparisons of automatic vs manual events of smartphone, automatic events vs ApneaLink, and manual events vs ApneaLink, respectively. Se and PPV are shown for each subject, as well as for the total number of episodes. The automatic algorithm performs similarly (Se=94.4%) than visual inspection (Table II A).

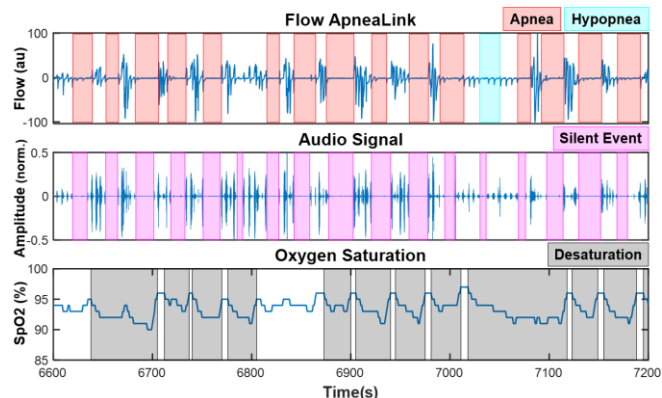


Figure 3. Example of a 10 minute segment from subject 2, comparing the ApneaLink flow channel (top) and its apnea (red) and hypopnea (cyan) labels, to smartphone audio signal (middle) and automatic silence event labels (magenta). Last channel is SpO<sub>2</sub> with desaturation labels (gray).

TABLE II. METHODS PERFORMANCE & COMPARISON

Subject	A) Automatic vs Manual		B) Automatic vs ApneaLink		C) Manual vs ApneaLink	
	Se	PPV	Se	PPV	Se	PPV
1	88 %	84.6 %	61.5 %	96 %	64.1 %	96.2 %
2	97.6 %	78.9 %	83.5 %	83.1 %	77.3 %	95.8 %
3	89.3 %	68.9 %	48.1 %	79.4 %	45.6 %	95.9 %
Total	94.4 %	77 %	69.3 %	83.4 %	65.4 %	95.9 %

When compared to ApneaLink, which is the reference of this study, automatic detection slightly outperforms manual labeling sensitivity to actual apnea or hypopnea events, but not PPV (Table II B-C). However, both Se values are lower than 70%. To further investigate this, we computed the percentage of apneas and the percentage of hypopneas that we were labeling as silence events. Results, shown in Table III, indicate that both methods are very sensitive to apneas, but miss about half the hypopneas. This is a limitation of the system which cannot be attributed to the automatic algorithm, but to the sound characteristics of hypopnea events.

Finally, Table IV shows the total number of events and the AHI estimated by each method. Smartphone AHIs tend to underestimate patients’ severity, especially those from manual procedure.

### IV. DISCUSSION

#### A. Manually Labeling Smartphone Signals

Medical labeling of smartphone signals was done to confirm if their quality and clinical value are enough to detect apnea and hypopnea events. We compared manual annotation to both automatic and ApneaLink labels. This helped us to assess whether differences between the automatic method and ApneaLink are more likely to be attributed to the algorithm or to the system (smartphone signals) itself.

Results indicate a similar performance of automatic and manual labeling to detect silence events. This suggests that our algorithm is correctly identifying silence regions, at least with an accuracy comparable to human eye. Moreover, more than 95% of manual labels (83% of automatic events) correspond to actual apneas or hypopneas (Table II). Therefore, these silence events have a physiological sense.

The manual approach finds a very low number of events (Table IV), which seems inconsistent with the sensitivity values in Table II. The explanation is that manually labeled events were much longer than the others. When comparing methods, if 2 events were present in one channel, but corresponded to just 1 event in the other or vice versa, they were counted as correct, so Se and PPV were not affected. However, this low number of events leads to an underestimated AHI (Table IV), thus highlighting the importance of correctly defining the event edges. An automatic detector can help to do that, since it is based on objective criteria. A serious limitation of AHI also arises: it just counts the number of events, ignoring their lengths. It will have the same value whether all the events have a duration of 10 s or 60 s, but the associated hypoxia and cardiovascular effects will be far more critical in the second case. For this reason, a new index could be preferred, e.g. the ratio of the sum of event lengths and the total signal duration.

TABLE III. PERCENTAGE OF APNEAS & HYPOPNEAS DETECTED

Subject	Automatic Detector		Manual Labeling	
	Apneas	Hypopneas	Apneas	Hypopneas
1	76.5 %	50 %	94.1 %	40.9 %
2	87 %	59.1 %	79.9 %	59.1 %
3	92.3 %	41.8 %	84.6 %	40 %
Total	86.2 %	46 %	81.5 %	44 %

TABLE IV. NUMBER OF EVENTS &amp; ESTIMATED AHI OF EACH METHOD

Subj	ApneaLink				Smartphone			
	Apn <sup>a</sup>	Hyp <sup>b</sup>	Events	AHI	Automatic		Manual	
					Events	AHI	Events	AHI
1	17	22	39	17.2	27	11.9	25	11
2	154	22	176	34.2	184	35.7	126	24.5
3	13	90	103	24.2	70	16.5	48	11.3

a. Number of Apneas / b. Number of Hypopneas

### B. Automatic Silence Events Detector

This work applies sample entropy to audio signals as a novel method to detect apnea/hypopnea episodes. A major challenge when working with acoustic breathing signals recorded with ambient microphones is the low signal-to-noise ratio. The main advantage of SampEn is its robustness and resistance against impulsive noise. This is crucial for us since, to correctly identify silence events (apneas or hypopneas), it is essential to reduce the effect of small impulsive artifacts, such as those from cardiac noise, swallows or upper airways and mouth “clicks”.

On the other hand, the main drawback of SampEn is its high computational cost. For this reason, we used the SpO<sub>2</sub> signal to guide our detector, and just analyze those regions followed by desaturations. Sleep technicians also use SpO<sub>2</sub> information when diagnosing patients, as most events related to sleep apnea condition are followed by desaturations (in fact, it is implicit in hypopneas definition). This approach reduces false alarms, at the cost of missing the apneas not followed by desaturations (although they are uncommon). Differences in pulse oximeters dynamics (ApneaLink vs Kenek Edge) can also produce detection errors.

Our automatic detector performed as well as manual labeling by a sleep expert, or even outperformed it. Se is higher with the automatic method (69% vs 65%), and a higher PPV is obtained when comparing it to ApneaLink than to manual labeling (83% vs 77%) (Table II). OSA and other diseases are usually still diagnosed by manual event labeling, but this is a tedious, high-cost and time-consuming task. We demonstrate that an automatic algorithm from audio signals may be used to screen sleep disorders, or at least to help the specialists and reduce their workload, improving the diagnostic accuracy and facilitating clinical decisions.

### C. Patients' Stratification: The Problem of Hypopneas

Automatic and manual detectors from smartphone signals are very sensitive to apneas but miss nearly half hypopneas. A higher sensitivity to apneas was expected, since hypopneas definition is complex and remains controversial. Nonetheless, this limitation of our mHealth system suggests that a certain acoustic activity remains during some hypopneas. Although hypopneas are defined as a flow reduction, it is not always reflected in a sound reduction. Future work is required to study this phenomenon. Even so, our automatic detector has great potential for non-invasive OSA monitoring, as it accurately identifies silence events. These are probably the most critical events, including apneas, and “silence hypopneas” (the most similar to apneas). This issue should be further explored, but these events carry interesting information for patients' identification, stratification and follow-up.

The problem of hypopneas reduces the sensitivity to ApneaLink events (Table II). This translates into an underestimation of AHI in those subjects with a higher percentage of hypopneas (Table IV). Predicted AHIs misclassify patient 1 as mild instead of moderate OSA. The same happens in patient 3 with manual inspection. Despite these inaccuracies, smartphone signals agree with ApneaLink indicating that subject 2 is the most severe, followed by subject 3. Increasing the sample size would help to understand the relationship between actual AHI and the index of silence events per hour.

## V. CONCLUSION

This work explores audio signals recorded with a smartphone microphone to extract information related to sleep apnea condition. We developed an algorithm to automatically detect silence episodes. It demonstrated a high accuracy, comparable to that of a sleep expert. Despite missing many hypopneas, it was quite sensitive to apneas.

This is just a pilot study, and the sample should be increased to optimize the algorithm and extract more robust conclusions. However, it is a proof-of-concept of the feasibility of mHealth systems and audio signal processing as a simple, low-cost and fast way to monitor sleep apnea patients at home and reduce the high impact of this disease.

## REFERENCES

- [1] H. Jin et al., “Acoustic analysis of snoring in the diagnosis of obstructive sleep apnea syndrome: a call for more rigorous studies”, *J Clin Sleep Med*, vol. 11, no.7, pp. 765–771, Jul 2015
- [2] C. V. Senaratna et al., “Prevalence of obstructive sleep apnea in the general population: A systematic review”, *Sleep Med Rev*, vol. 34, pp. 70-81, Aug 2017
- [3] American Academy of Sleep Medicine (AASM), “The AASM Manual for the Scoring of Sleep and Associated Events v2.5”, 2018
- [4] R. Jané, J. A. Fiz, J. Solà-Soler, J. Mesquita and J. Morera, “Snoring analysis for the screening of sleep apnea hypopnea syndrome with a single-channel device developed using polysomnographic and snoring databases”, *Conf Proc IEEE Eng Med Biol Soc*, pp. 8331-8333, Aug 2011
- [5] J. Behar et al., “SleepAp: an automated obstructive sleep apnoea screening application for smartphones”, *IEEE J Biomed Health*, vol. 19, no. 1, pp. 325-331, 2015
- [6] M. Al-Mardini, F. Aloul, A. Sagahyoon and L. Al-Husseini, “Classifying obstructive sleep apnea using smartphones”, *J Biomed Health*, vol. 52, pp. 251-259, 2014
- [7] M. P. Bonnesen, H. B. D. Sorensen, and P. Jennum, “Mobile apnea screening system for at-home recording and analysis of sleep apnea severity”, *Conf Proc IEEE Eng Med Biol Soc*, pp. 457-460, Jul 2018
- [8] Y. Castillo, M. A. Cámara, D. Blanco-Almazán and R. Jané, “Characterization of microphones for snoring and breathing events analysis in mHealth”, *Conf Proc IEEE Eng Med Biol Soc*, pp. 1547-1550, Jul 2017
- [9] M. A. Cámara, Y. Castillo, D. Blanco-Almazán, L. Estrada and R. Jané, “mHealth tools for monitoring obstructive sleep apnea patients at home: proof-of-concept”, *Conf Proc IEEE Eng Med Biol Soc*, pp. 1555-1558, Jul 2017
- [10] H. Nakano et al., “Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept”, *J Clin Sleep Med*, vol.10, no.0.1, pp. 73-78, 2014
- [11] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy”, *Amer J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2030-H2049, 2000
- [12] L. Estrada, A. Torres, L. Sarlabous and R. Jané, “Improvement in neural respiratory drive estimation from diaphragm electromyographic signals using fixed sample entropy”, *IEEE J Biomed Health*, vol. 20, no. 2, pp. 476-485, Mar 2016