

# scientific data



OPEN

DATA DESCRIPTOR

## PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies

Georgia Korompili<sup>1</sup>, Anastasia Amfilochiou<sup>2</sup>, Lampros Kokkalas<sup>1</sup>, Stelios A. Mitilineos<sup>1</sup>, Nicolas- Alexander Tatlas<sup>1</sup>, Marios Kouvaras<sup>1</sup>, Emmanouil Kastanakis<sup>2</sup>, Chrysoula Maniou<sup>2</sup> & Stelios M. Potirakis<sup>1</sup>✉

The sleep apnea syndrome is a chronic condition that affects the quality of life and increases the risk of severe health conditions such as cardiovascular diseases. However, the prevalence of the syndrome in the general population is considered to be heavily underestimated due to the restricted number of people seeking diagnosis, with the leading cause for this being the inconvenience of the current reference standard for apnea diagnosis: Polysomnography. To enhance patients' awareness of the syndrome, a great endeavour is conducted in the literature. Various home-based apnea detection systems are being developed, profiting from information in a restricted set of polysomnography signals. In particular, breathing sound has been proven highly effective in detecting apneic events during sleep. The development of accurate systems requires multitudinous datasets of audio recordings and polysomnograms. In this work, we provide the first open access dataset, comprising 212 polysomnograms along with synchronized high-quality tracheal and ambient microphone recordings. We envision this dataset to be widely used for the development of home-based apnea detection techniques and frameworks.

### Background & Summary

The sleep apnea syndrome (SAS) is a breathing disorder occurring as a repetitive cessation or severe reduction of breathing, leading to disrupted sleep<sup>1,2</sup>. Although it was noticed even before 1900, under the term "Pickwickian syndrome", in exceptional cases linked to obesity and hypersomnolence, SAS was recognized as a pathological chronic condition only after 1970<sup>3</sup>. Since then, various studies attempted to determine the SAS relationship with age<sup>4–8</sup> or gender<sup>4,7,9–12</sup> and with pathological conditions like obesity<sup>7,13–17</sup>, hypertension<sup>16,18–21</sup>, diabetes mellitus<sup>22–25</sup>, previously reported stroke<sup>26–29</sup> or cardiovascular diseases<sup>30–34</sup> and cancer<sup>35</sup>.

The SAS symptoms include daytime fatigue and sleepiness<sup>36,37</sup>, headaches, depression or mood changes<sup>37,38</sup>, disrupted interpersonal relationships<sup>38</sup>, reduced cognitive performance<sup>37,38</sup> and increased risk of work and vehicle accidents<sup>39–41</sup>. Despite the symptoms' severity, the awareness of the syndrome remains restricted<sup>42</sup>. The epidemiological studies on apnea prevalence in the general population exhibit increased inconsistency (3–7% for men, 2–5% for women)<sup>43–45</sup>. This is mainly attributed to the medical protocols' differences<sup>46</sup>. However, the restricted number of people seeking diagnosis also strongly contributes to SAS prevalence underestimation<sup>47</sup>.

The reference standard for diagnosing apnea is Polysomnography (PSG): a study of cardiac, neurological and respiratory signals over a full night sleep in hospital<sup>48</sup>. The PSG inconvenience, along with the requirement for full night hospitalization, discourages patients from seeking diagnosis. Thus, multiple systems for SAS estimation at home have been developed. They provide easier examination and increase patients' awareness of SAS. The reported studies profit from smartphones' ubiquity, their connectivity to portable sensors and their computational power, which allows for data processing through built-in applications<sup>49–55</sup>. Breathing sound has been proven highly effective in this endeavour<sup>56,57</sup>. Most studies employ neural networks (NNs)<sup>58–62</sup>, thus, they strongly depend on annotated PSG data and audio recordings via ambient<sup>49</sup> or contact microphones<sup>59,60</sup>.

<sup>1</sup>Department of Electrical and Electronic Engineering, University of West Attica, Attica, Greece. <sup>2</sup>Sleep Study Unit, Sismanoglio – Amalia Fleming General Hospital of Athens, Athens, Greece. ✉e-mail: [spoti@uniwa.gr](mailto:spoti@uniwa.gr)

The data collection process is crucial for the effectiveness and accuracy of the evolved systems, contributing to both the systems' development and validation. But data collection is a time-consuming process, requiring a large number of patients to undergo a PSG in hospital. Online available datasets are a convenient alternative to save valuable time and assure feasibility of comparison between different approaches<sup>63–67</sup>. However, to the best of our knowledge, there is currently no available dataset that includes breathing sound recordings simultaneous to PSG. Raw PSG data without contact or ambient microphone recordings are publicly available mainly through the National Sleep Research Resource<sup>65</sup>, referring to sleep studies on the general population<sup>65,68,69</sup> or on specific sub-populations such as paediatric patients<sup>70,71</sup>, pregnant women<sup>72</sup> or elderly men<sup>73,74</sup>. While tracheal microphones participate in standard PSG systems, the recordings are usually of poor quality<sup>75</sup> with low sampling frequency or narrow dynamic range. The lack of standardization protocols with regard to the microphone type, bandwidth and signal compression leads to inconsistent results between the developed methodologies. Even during the implementation of proprietary datasets, sound is rarely recorded simultaneously with PSG, with an exception being the study of Azarbarzin and Moussavi<sup>76</sup>. In audio-oriented studies, sound is usually recorded separately<sup>77</sup> and reference scoring of apneas relies solely on sound or a restricted number of PSG signals<sup>78</sup>, rendering diagnosis incomparable to PSG findings.

To meet these challenges, we built and provide an open access dataset comprising 212 PSGs with synchronized sound recordings from tracheal and ambient microphones. The data are collected and characterized by the medical team of Sismanoglio – Amalia Fleming General Hospital of Athens, and are open and freely available online [<https://doi.org/10.11922/scientedb.00345>]<sup>79</sup>. We trust that this dataset will particularly contribute in: (a) comparable studies on the effectiveness of contact and ambient microphones in apnea detection, (b) studies on breathing sound features for specific time detection of apneic episodes and (c) data augmentation of PSG-related studies. The dataset is part of an on-going research study and is expected to be enriched in the following months/years.

Herein, we present the data collection methodology, the emerging ethical issues management, as well as several key dataset statistics since a complete characterization of an open access dataset is highly important in order to support future research findings<sup>80</sup>. A discussion is additionally conducted concerning crucial challenges of this field, particularly the subjectivity of manual respiratory events labelling and the requirements for multitudinous and balanced datasets for developing accurate SAS detection systems.

## Methods

**Ethical issues management.** The participation of the patients in the present study is approved by the Local ethics committee of Sismanoglio Hospital and gives rise in several ethical issues that need to be managed following the European Regulation for Personal Data Protection<sup>81</sup>. All patients were asked to give signed consent for participating in the study and agree in audio signal recording during their sleep. They were also asked to agree on the use of all unnamed recorded data for research purposes. The involvement of the health care personnel in the process of the PSG medical examination, providing instruction and help to the patient, requires additional consent for their speech recording at the beginning or during the study.

Data acquired from each patient were stored in the hospital and processed by the health specialists to extract diagnosis. All personal information leading to identification of the participating individuals (names, credentials, contact information, etc.) was removed from the acquired files. An unnamed copy of each PSG study along with the corresponding audio files were stored and further processed to participate in the dataset.

**Signal acquisition and storage.** The data were collected from 212 individuals, who visited the Sleep Study Unit of the Sismanoglio – Amalia Fleming General Hospital of Athens for SAS diagnosis. The patients were subjected to a full night sleep study following the standard protocol for split PSG<sup>82</sup> in which the first part – approximately four (4) hours – is a standard diagnostic PSG while the second part is used for titration to optimal level of pressure, to eliminate apnea events with continuous positive airway pressure (CPAP). The optimum level pressure for all patients following split study protocol is also available in the dataset. If the patient was not suitable for split study, diagnostic PSG was not interrupted. Since the developed dataset is built for studying sleep without CPAP intervention, only the first part of the recorded study was included. A detailed listing of the signals, recorded through the PSG system and the respective sampling frequency for each channel are given in Table 1. PSG channels monitoring and signal acquisition was performed using the Sleepware G3 software.

Simultaneously with the PSG study a dual channel portable multitrack recorder (Tascam DR-680 MK II) was used in order to acquire and store the audio signals from two high-quality microphones: (a) a contact microphone (Clockaudio CTH100) placed on the trachea of the patient and (b) an ultra-linear measurement condenser microphone (Behringer ECM8000) placed approximately 1 m above the patient's bed, over the head position (Fig. 1). Both sound signals are sampled at 48 kHz and stored in an SD card as 24-bit uncompressed Waveform Audio Format (.wav) files. The contact electret microphone (input impedance: 900 Ω, passband: 350 Hz – 8 kHz) acquires only the neck vibrations while it is completely insensitive to environmental noise. The ambient condenser microphone is of electret type, omnidirectional, with an input impedance of 200 Ω and a flat frequency response in the range of 15 Hz–20 kHz.

**Medical data annotation process and diagnosis extraction.** Medical characterization of each particular PSG study is performed by the health specialists of the Sleep Study Unit of the Sismanoglio – Amalia Fleming General Hospital of Athens. For each patient, sleep stages and apnea events are scored by two specialists: a certified technician performs first level scoring and a 30-year-experienced and certified doctor performs final scoring, with verification of the true positive annotated events and addition of missed events. The process of scoring between the specialists is not blind, however, the followed protocol assures increased accuracy. Inter-observer agreement was evaluated in the past, maximizing the provided accuracy through the followed process, while further examination of the inter- and intra- observer error is beyond the goals of the current project. Additionally,

Channel label	Description	Sampling frequency (Hz)	Physical minimum (a.u.)	Physical maximum (a.u.)	Digital minimum	Digital maximum
'EEG A1-A2'	Electroencephalogram (reference channel)	200	-313	313	-32768	32767
'EEG C3-A2'	Electroencephalogram (channel 1)	200	-313	313	-32768	32767
'EEG C4-A1'	Electroencephalogram (channel 2)	200	-313	313	-32768	32767
'EOG LOC-A2'	Electroculogram (left eye)	200	-313	313	-32768	32767
'EOG ROC-A2'	Electroculogram (right eye)	200	-313	313	-32768	32767
'EMG Chin'	Electromyogram (placed on chin)	200	-78	78	-32768	32767
'Leg 1'	Leg movement signal (right leg)	200	-78	78	-32768	32767
'Leg 2'	Leg movement signal (left leg)	200	-78	78	-32768	32767
'ECG I'	Electrocardiogram	200	-8333	8333	-32768	32767
'RR'	RR interval in the ECG	10	0	200	0	200
'Snore'	Contact microphone (placed on trachea)	500	-100	100	-32768	32767
'Flow Patient'	Thermistor	100	-100	100	-32768	32767
'Flow Patient'	Pressure cannula	100	-100	100	-32768	32767
'Effort THO'	Respiratory belt (thoracic volume changes)	100	-100	100	-32768	32767
'Effort ABD'	Respiratory belt (abdomen volume changes)	100	-100	100	-32768	32767
'SpO <sub>2</sub> '	Oxymeter	1	0	102.3	0	1023
'Body'	Position of the body (S = supine, R = right, L = left)	1	0	255	0	255
'PulseRate'	The pulse rate extracted by the ECG	1	0	255	0	255
'Tracheal'	High-quality contact microphone (placed on trachea)	48000	-100	100	-32768	32767
'Microphone'	Ambient microphone (placed ~1 m above patient head)	48000	-100	100	-32768	32767

**Table 1.** Basic properties of the channels included in the EDF files of the dataset.**Fig. 1** Contact and ambient microphones installed in the sleep study room, along with the multitrack recorder.

we developed an algorithm that quantifies the decrease in the flow rate amplitude within each scored event to validate data and assure high accuracy in the true positive events annotation. The algorithm details and the results are presented in Section 3.5.

The scoring of sleep stages during the total sleep time (TST) relies on the general instructions for sleep stage labelling<sup>1,83</sup>. The detection of apnea/hypopnea events, during the recorded sleeping hours, was performed manually by simultaneous observation of all channels of the PSG system, according to the general criteria for apnea episode scoring. Audio recordings were not included in the diagnostic process. The final diagnosis concerning the categorization of the patient in one of the reported apnea severity cases: "Severe", "Moderate", "Mild Apnea" and "Normal" was extracted through the Apnea/Hypopnea Index (AHI). The AHI is defined as the ratio of the total count of apneic episodes in the entire sleep study over the TST in hours [1], which results in the mean count of apneic events per sleeping hour. Up to 5 apnea/hypopnea episodes per hour classify the subject in the case of "Normal breathing" during sleep while higher values indicate a gradually increasing severity of SAS ( $5 \text{ episodes/h} \leq \text{AHI} < 15 \text{ episodes/h}$ : "Mild Apnea",  $15 \text{ episodes/h} \leq \text{AHI} < 30 \text{ episodes/h}$ : "Moderate Apnea",  $30 \text{ episodes/h} \leq \text{AHI}$ : "Severe Apnea")<sup>1,2</sup>.

**Signals synchronization.** A critical step is the synchronization of the acquired audio signals with the PSG signals derived from separately activated systems. The synchronization algorithm employs the tracheal sound

Event Family	Specific types
Respiratory	'Obstructive Apnea'/'Central Apnea'/'Mixed Apnea'/'Hypopnea'/'Cheyne Stokes Respiration'/'Periodic Respiration'/'Respiratory effort- related arousal (RERA)'
Neurological	'Arousal'
Limb	'Leg Movement'/'Alternating Leg Muscle Activation'/'Hypnagogic Foot Tremor'/'Excessive Freamgmentary Myoclonus'/'Rhythmic Movement Disorder'
Nasal	'Snore'
Cardiac	'Bradycardia'/'Tachycardia'/'Long RR'/'Ptt Drop'/'Heart rate Drop'/'Heart Rate Rise'/'Asystole'/'Sinus Tachycardia'/'Narrow Complex Tachycardia'/'Wide Complex Tachycardia'/'Atrial Fibrillation'
SpO <sub>2</sub>	'Relative Desaturation'/'Absolute Desaturation'

**Table 2.** Annotated events family and corresponding types.

recorded from a separate, low quality, contact microphone of the PSG system (channel label "Snore") – sampling frequency: 500 Hz and bit depth: 16-bits. The two acquired tracheal sound signals were synchronized by extracting the signal envelope with Hilbert transform. Prior to this step, it was necessary to reduce the sampling rate of the high-quality tracheal signal so that the two signals have a common sampling frequency of 500 Hz. The cross correlation of the two signal envelopes is extracted and the delay in the activation of the two systems is estimated by the time difference value that maximizes cross correlation. To accurately determine this delay, we examine at least 10 min of the signals, while the supervisors of the sleep study are given the instruction to activate the two systems with the minimum possible delay, most frequently below 30 s. Manual observation of the signals results in the estimation of the error in synchronization and the rejection of those PSG studies that exhibit an error higher than 2 s in this step. Thus, among a total number of 240 patients that underwent PSG study between April 2019 and July 2020, 28 patients were rejected from the herein presented dataset as they exhibit low quality audio or PSG recordings – mainly due to the dislocation of one or more sensors – and the corresponding signals could not be accurately synchronized.

### Data Records

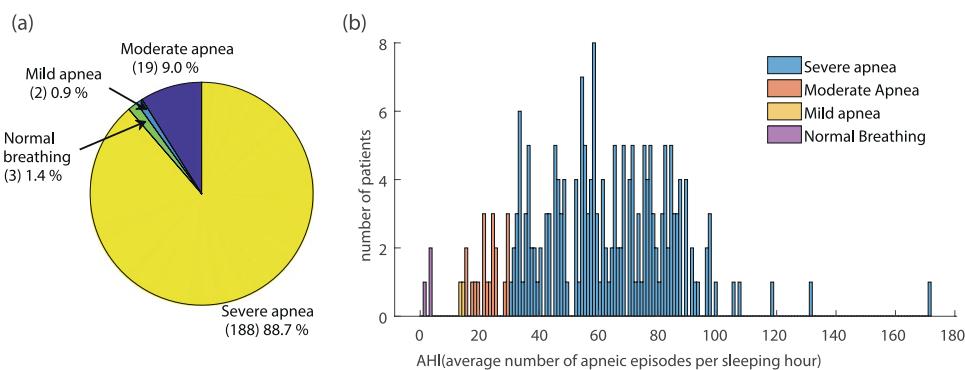
The storage of polysomnographic and breathing sound data was performed by employing the European Data Format (EDF) common in medical data storage and transfer<sup>84</sup>. The EDF files contain the channels listed in Table 1, with the corresponding sampling frequency and digital and physical maximum and minimum values. While the PSG data were retrieved from equivalent EDF files and remain unchanged, the additional audio data were retrieved from uncompressed WAV files and stored in the EDF files under the channel names "Tracheal" and "Microphone", corresponding to the tracheal sound signal and the ambient microphone signal respectively. The import of the separately recorded high-quality audio data ("Tracheal" and "Microphone") was performed by using zero padding policy for the missing parts of the signals, as a result of the synchronization process. EDF is selected due to its popularity in medical data storage and transform, despite the fact that, for the audio signal, this format results in a reduction of the bit depth. Indeed, the most common version of EDF files requires storage in 16-bit while the original WAV files were stored in 24-bit depth. Alternatively, BioSemi Data Format (BDF) files could be used, which is a 24-bit version of EDF files; however, this option was rejected due to its still restricted popularity and the existence of fewer software platforms that support it. The conversion of the initially acquired high-quality audio recordings from 24 to 16-bit depth is performed by neglecting the least significant bits of each sample value. The statistical absolute error due to this conversion presents an average value of  $7.6291 \cdot 10^{-6}$  while the maximum value does not exceed  $1.5259 \cdot 10^{-5}$ . These values impose minor quality reduction in the audio recordings that is not expected to affect the studies related to breathing or snoring sound properties. The EDF files corresponding to each patient study are cropped in parts of 1 h duration each, to facilitate handling. Each patient was labelled with a unique representative patient number (range 993–1496).

In the dataset, additional .rml files are responsible for all annotations corresponding to each patient. These files include the sleeping stages and the labelled events. The annotated events are characterized by the family in which they belong ("respiratory", "neurological", "limb activity related", "nasal" and "cardiac") and the type, which is related to each event family according to Table 2. In particular, the "respiratory" episodes, which are the main concern of this report, include among others all apnea related episodes of specific type: "Obstructive Apnea", "Central Apnea", "Mixed Apnea" and "Hypopnea". Additionally, the.rml files contain all annotated episodes of relative oxygen desaturation and arousal events. Concerning the patients' data, the information on the age and gender of each participating individual were kept in the corresponding.rml files.

All data are stored in an open access dataset, available for free download here: <https://doi.org/10.11922/sciencedb.00345><sup>79</sup>.

### Technical Validation

**Basic statistical features on the divergence of patients participating in the dataset.** A major issue in the use of polysomnographic data in related studies and systems development is the balancing between different categories of patients that are included in the used dataset. The divergence of the participating subjects with regards to factors such as the gender, the age and the final diagnosis of SAS severity may noticeably alter the features of recorded breathing and snoring sound and the episodes properties such as the duration of each episode. The PSG is usually prescribed to patients who complain about excessive sleepiness during daytime or loud snoring during sleep, symptoms that are strongly related to the presence of SAS. As expected, the majority of the participants (88.7%) belong to the group of "Severe Apnea" while the percentage of "normal" cases is restricted,



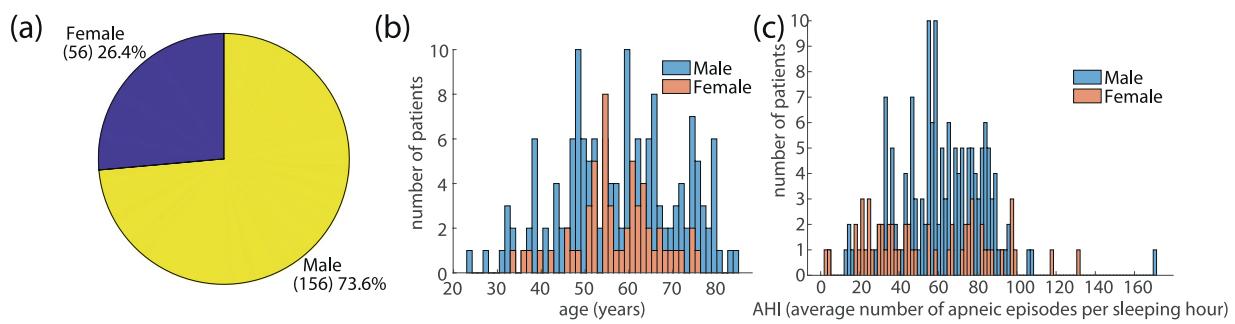
**Fig. 2** Dataset statistics on patients' apnea syndrome severity. **(a)** Patients' distribution among the classes of different apnea severity. **(b)** The distribution of AHI value in all subjects participating in the dataset with labeling of the corresponding apnea severity class.

not exceeding 1.4% of all diagnosed individuals (Fig. 2a). Taking into account the increased risk of SAS in male population, the gender classification of the participating subjects is strongly imbalanced, with male population representing 76% of the entire dataset (Fig. 3a). The age distribution of patients ranges from 34 to 76 years for women and 23 to 85 years for men. The mean values are 57.2 and 57.9 years for women and men respectively, with different ages equally distributed in men and women (Fig. 3b).

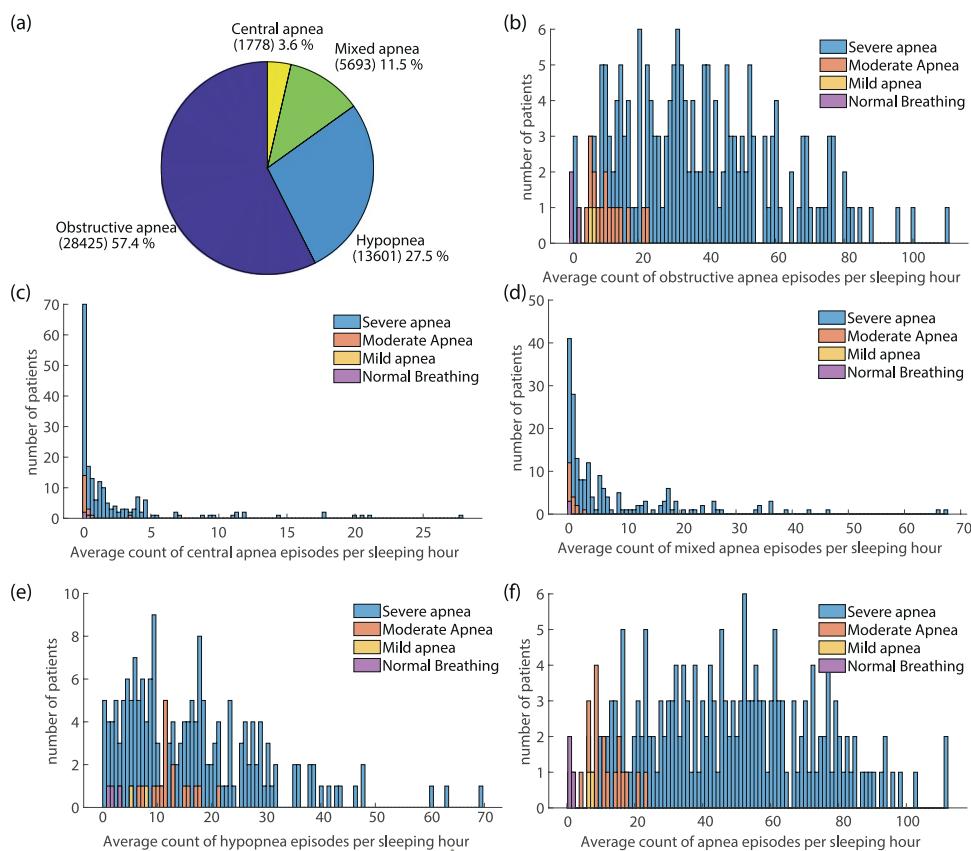
The age distribution of the total population participating in the dataset exhibits significant similarities between the different groups of apnea severity classes (see Supplementary Fig. 1) despite the existence of imbalanced data among them. It is also interesting that the distribution of AHI extends over approximately the same range for both male and female individuals (Fig. 3c). These statistics assure that although the dataset is subjected to severe unbalancing between the different groups, the dataset information concerning apnea/hypopnea episodes covers a wide range of AHI – consequently all SAS severity classes – and a wide range of ages. It is reminded that all statistical measures presented here should not be considered as epidemiological data but only as features indicative of the dataset's balancing.

**The annotated apnea related episodes.** The labelled apnea/hypopnea events belong to the “Respiratory” family and were further subcategorized in the corresponding types of “Obstructive Apnea”, “Central Apnea”, “Mixed Apnea” and “Hypopnea”. Although the software in use for manual observation and annotation of data (Sleepware G3) allows for labelling of subtypes of hypopnea, the latest protocol for apnea scoring<sup>1</sup>, followed in this work, suggests the subdivision of hypopnea events into the cause-related types (Obstructive and Central) to be avoided<sup>85,86</sup>. The majority (99.94%) of all annotated respiratory events (total sum 49525 episodes) were specific types of Apnea (35896 events) or Hypopnea (13601 episodes). As expected, the obstructive sleep apnea episodes dominate in frequency of appearance among all annotated apneic episodes (57.4%), while central apnea events represent only 3.6% of the total count of labelled episodes (Fig. 4a).

The criteria for labelling all types of different apnea episodes are clearly defined in the protocol for sleep apnea scoring<sup>1</sup>. These criteria refer to air flow signals – measured through pressure drop and air thermal changes close to patient's nose – or to the thoracoabdominal movements representing the breathing effort. The oxygen relative desaturation and possible arousal – indicated by the corresponding neurological signals – are employed as additional factors contributing to the safe identification of apneic episodes. The employed criteria are equal for all patients. The sound signal does not participate in the diagnosis process. However, the research approaches relying on snoring and breathing sound recordings to perform apnea detection, frequently discuss the variation of sound features between patients with different SAS severity. An apneic episode may significantly differ in terms of sound characteristics for a mild snorer compared to a heavy snorer, while snoring characteristics are strongly related to the presence of SAS. Therefore, the distribution of annotated episodes in patients belonging to a different SAS severity class is considered an important factor, indicative of the expected variability of sound features. Figure 4(b–e) summarizes the distribution of the main four categories of apneic events (“Obstructive”, “Central”, “Mixed” and “Hypopnea”) per sleeping hour in relation to the patients' overall diagnosis (SAS severity class: “Severe”, “Moderate”, “Mild” and “Normal”). As expected by the prevalence of central apnea syndrome, the frequency of central and mixed apnea episodes is concentrated in the range between 0–5 and 0–20 apneas per sleeping hour, respectively, independently of the severity group in which the patient is classified. The obstructive apnea events represent the major contributor to the final AHI and consequently to the overall diagnosis for the patient<sup>87,88</sup>, with the variation in the frequency of appearance resembling the AHI distribution. The hypopnea events of this dataset are particularly frequent in the case of moderately apneic individuals, though such a deduction should also take into account the restricted number of moderate apneic subjects participating in the dataset and it requires further investigation. In Fig. 4e,f a comparison of the Apnea Index (AI) and Hypopnea Index (HI) is attempted. The AI is extracted as the sum of all types of apnea events per sleeping hour, while HI includes only the hypopnea episodes present per hour of sleep. Except for the recent tendency towards separate interpretation of these indices for more accurate diagnosis of SAS severity, the aforementioned comparison clearly indicates that apnea events contribute mostly to the overall SAS severity estimation for this dataset. This might be a useful aspect to consider in the development of sound-based systems for SAS detection, taking into account that the air



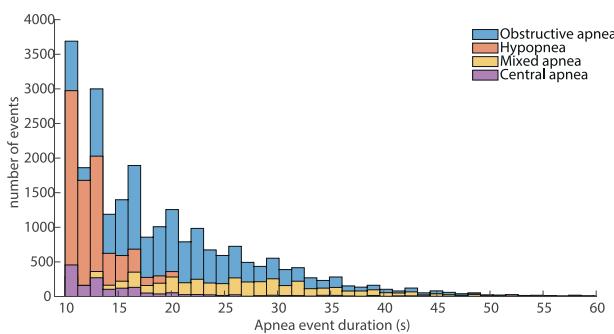
**Fig. 3** Dataset variability with regards to the patients' age and gender. **(a)** Gender distribution and **(b)** age distribution examined separately for men and women for all subjects participating in the dataset. **(c)** AHI distribution separately for the male and female participants.



**Fig. 4** Dataset variability with regards to the scored apnea/hypopnea episodes. **(a)** Frequency of appearance of different types of apnea/hypopnea episodes in the provided data. The distribution of **(b)** obstructive apnea, **(c)** central apnea **(d)** mixed apnea and **(e)** hypopnea events per sleeping hour for patients belonging in different SAS severity class. **(e)** and **(f)**: comparison of Apnea Index (AI) and Hypopnea Index (HI) as separate indicators contributing to the final overall characterization of the patient with regards to SAS severity.

flowing, through the completely or partially collapsed upper airway system of the patient respectively, must result in significantly different sound characteristics.

Additionally, Fig. 5 illustrates the distribution of episodes' duration with regard to the types of apnea/hypopnea. The duration of apnea/hypopnea episodes is believed to represent a particularly interesting factor in the process of sleep study interpretation, suggested by multiple researchers as a separate index to be measured. It is believed that the duration of an apneic episode highly correlates with the effects of relative oxygen desaturation and probably with significant hypoxemia<sup>89</sup>. In the provided dataset, the duration of labelled episodes ranges from 10 s (minimum duration of an acceptable apneic episode) to 128.5 s corresponding to a specific "mixed" apnea event. The distribution of apneic events duration, studied separately for the four main types of apnea, indicates that "mixed" apneas exhibit a significantly higher mean duration (24.6457 s) compared to all other types – 19.0656 s, 16.1016 s and 16.162 s mean duration for "obstructive", "central" and "hypopnea" events, respectively.



**Fig. 5** Distribution of the duration of events labeled in the dataset. The distribution is given separately for the main four categories of apnea/hypopnea events (“Obstructive”, “Central”, “Mixed” and “Hypopnea”).

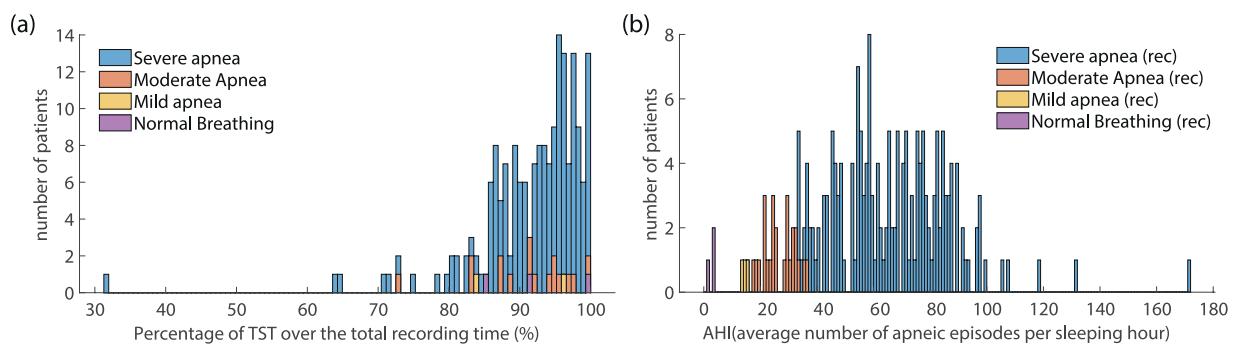
The distributions of the mean and maximum event duration per patient do not exhibit significant differences that could classify patients into the four SAS severity categories (see Supplementary Fig. 2) with the longer apneas being present in “Severe” apnea patients and the shorter in “Normal breathing” subjects.

**Sleeping vs. recording time and intra-night AHI variability in the collected PSGs.** The development of home-based systems aims at patients’ convenience; therefore, a restricted number of sensors are employed. A major argument questioning the provided accuracy by home-based systems in comparison to the gold standard of PSG study is the inability of these systems to count the TST of the patient. Assuming that the total recording time can replace the TST, home-based apnea detectors may result in a severe underestimation of the AHI. In the provided dataset, the TST of the patients – identified through the PSG neurological channels – is compared to the total recording time. For the majority of the participating individuals (95.05%) the TST was more than 80% of the recording time, while for approximately 70% of them the difference between the two values did not exceed 10% (Fig. 6a). These results are practically independent from the severity class.

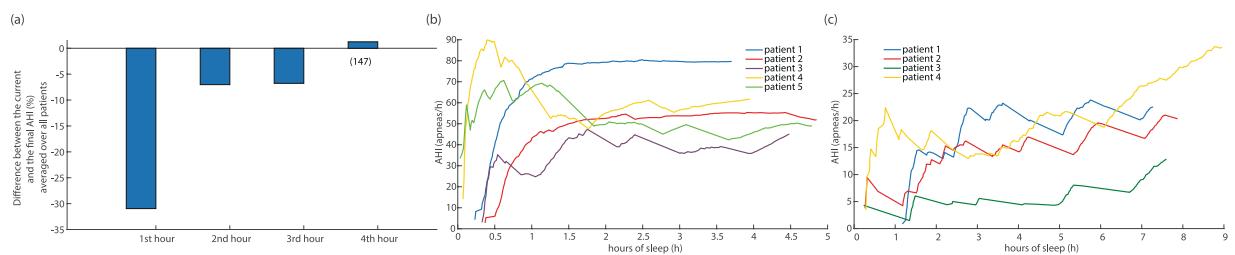
The impact on the AHI and SAS severity estimation can be calculated through the comparison of the AHI – derived by the TST of each patient – and the hypothetical SAS severity estimation based on the ratio of the number of apneas/hypopneas over the total recording time of each individual study. The results, illustrated in Fig. 6b, indicate that only 8 out of 212 patients would have been misdiagnosed. In particular, the change in classification appears mainly as a demotion from “Severe” to “Moderate” apnea classes.

In the same context, intra-night variability of the AHI can noticeably alter the final diagnosis in case only a few hours of sleep are taken into account. Recent publications, in the field of AHI interpretation, prove the significant variability of the index through the night and particularly the increase of obstructive apnea events frequency towards morning<sup>90</sup>. More precisely, the researchers suggest that the use of the highest AHI met through a time frame of 2 hours could be beneficial in reaching high level of accuracy in the SAS severity diagnosis. In the dataset provided herein, the intra-night variations of AHI are presented though (a) the study of the first 3 hours of sleep and (b) the difference between the current and the final AHI – derived as the averaged frequency of apneic events over the TST. In this particular statistical value, representing the dataset, we excluded the patients that did not complete a minimum duration of 3 h of recordings in the split-study protocol. Thus, only 177 patients appear in the plot of Fig. 7a. It is indicated that the mean difference between the current and total AHI (averaged for all subjects) gradually decreases reaching ~7% after the second hour of sleep (Fig. 7a). However, this difference, studied particularly for each patient, proves the existence of cases where the AHI of the first 3 hours of sleep is significantly lower than the final AHI (see Supplementary Fig. 3). By studying some specific cases of patients, we noticed the presence of a large number among them, in which the AHI remains stable after the 2nd sleeping hour, with the variations in these first 2 hours of sleep either overestimating or underestimating SAS severity (Fig. 7b). By studying the cases of patients subjected to a full night sleep protocol (total recording time at least equal to 7 h) we noticed that the variation of AHI may be important, leading to a final diagnosis change even within the last hour of sleep (Fig. 7c). It is therefore concluded that longer recordings can cover all different cases and lead to more accurate diagnosis, while it might be beneficial for home-based systems to rely on the recordings after the first two hours of sleep where the variations of AHI seem to reduce with respect to the final index.

**Additional PSG channels: the SpO<sub>2</sub> and arousal events used in apneic/hypopneic events’ detection.** Oxygen desaturation during sleep is of crucial importance since it indicates the degree of hypoxemia<sup>91,92</sup>. The Oxygen Desaturation Index (ODI) is defined as the average number of oxygen desaturation episodes per sleeping hour and it strongly correlates with the AHI of a patient<sup>93,94</sup>. In the dataset reported herein, the ODI exhibits an ascending behaviour when AHI increases (Fig. 8a). However, the indicated factors quantifying the goodness of fit with a linear relationship between the two indices (sum of squares due to error - SSE: 5.881e+04, R-square: 0.6801, Adjusted R-square: 0.6786, root mean squared error - RMSE: 16.73) prove that there are cases of patients exhibiting low ODI while their AHI classifies them in the upper groups of SAS severity (“severe” or “moderate”). Based on the literature, the ODI is often studied, either along with AHI or separately, for the diagnosis of SAS severity<sup>95</sup>. In Fig. 8b, we illustrate the ODI distribution for each class of SAS severity for all patients of this dataset. It is clearly depicted that the vast majority of patients characterized by severe apnea present an increased ODI while those with moderate or mild apnea exhibit a much lower ODI.



**Fig. 6** Statistics of the dataset with respect to the scored TST and the total recording time for each patient. (a) Difference between the TST and the total recording time, expressed as a percentage over the total recording hours. (b) The estimated AHI when the total recording time is taken into account instead of the TST, for each patient. The provided AHI should be used as reference value for the development of home-based systems that do not perform detection of the sleep stages.

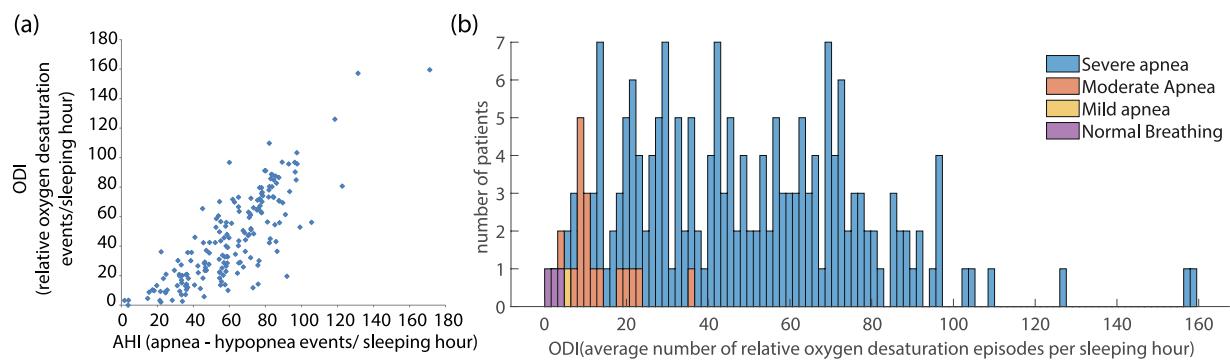


**Fig. 7** Dataset description with respect to the AHI intra-night variability. (a) The mean (averaged over all patients) difference between the current and final AHI for the first four hours of sleep. Only the patients with at least 3 complete hours of sleep (206 out of 212) participate in the extracted statistical values, while the average difference between the 4th hour AHI and the final AHI is extracted by 147 patients that have completed the 4th hour of sleep. (b) The intra-night variation of AHI for 5 patients following a split PSG study protocol. (c) The intra-night variation of AHI for 4 patients that have completed at least 7 hours of recordings without CPAP.

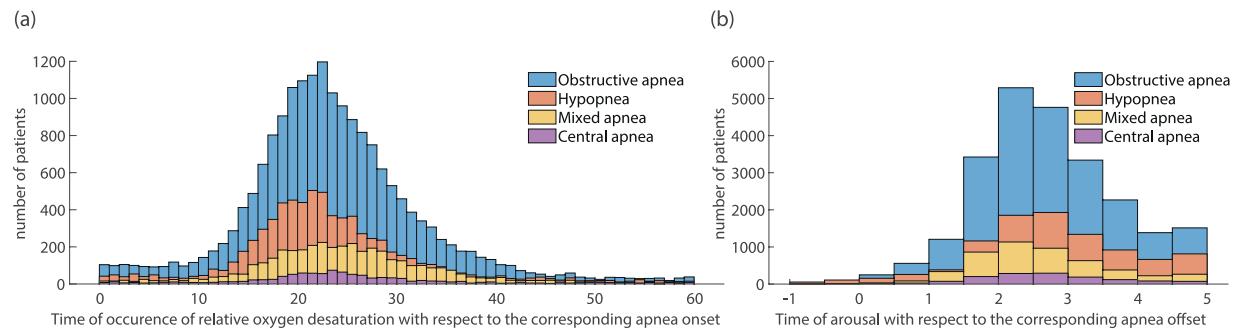
The association of ODI with AHI is partially explained by the fact that an apneic episode is frequently (but not always) followed by an episode of relative decrease of oxygen saturation; an oxygen desaturation event is defined as an episode of at least 3% reduction in the oxygen saturation level. In many available systems, these events have been employed as indicative of the preceding apneic event<sup>96</sup>. Herein, we report basic statistical values of the oxygen desaturation events that are linked to a specific event of apnea/hypopnea (of any type: “Obstructive”, “Central”, “Mixed apnea” or “Hypopnea”). The criterion for the association of the two events is not clearly determined in the literature. Some studies have used the criterion that the oxygen desaturation occurs within a frame of 60 s with reference to the onset of the associated apnea/hypopnea episode<sup>97</sup>. We follow the same criterion resulting in 31178 apnea/hypopnea events – among the sum of 49497 events annotated in this dataset – followed by a relative oxygen desaturation episode, thus a percentage of 62.99% of all labelled events. Particularly, 65.63% of the “obstructive” apneas, 60.24% of “central” apneas and 70.88% of “mixed” apneas were followed by a desaturation event. A lower percentage of “hypopneas” is accompanied by an episode of oxygen saturation drop (only 54.53%).

Among the SpO<sub>2</sub> episodes that are associated with a specific apnea/hypopnea event, 94.5% appear with a delay of at least 10 s with reference to the onset of the event. From the distribution of the time delay in the occurrence of oxygen desaturation in the different types of apneic events it is clear that negligible differences appear between the different types of apnea/hypopnea (Fig. 9a). Moreover, for this dataset the mean delay in an SpO<sub>2</sub> event occurrence does not seem to correlate with the SAS severity class (see Supplementary Fig. 4), despite what is recently discussed in literature<sup>97</sup>. The statistical features presented in this section should be seen through the prism of the imbalanced number of subjects participating in each SAS severity group but also of the fact that the manual process for apneic events labelling was performed independently from the oxygen desaturation events’ scoring.

In the same context, the arousals associated with the presence of apnea/hypopnea events have been studied. Regarded as an immediate consequence of an apnea, an arousal accompanies an apneic event and results in cessation of sleep. In this dataset, a percentage of 80.9% (40042 events) of all annotated apneas – of any type – are followed by an immediate arousal event, occurring within a time frame of 5 s after the end of the apnea/hypopnea episode. When studying separately each type of apnea: “obstructive”, “mixed” and “central” apneas present a percentage of 84.83%, 86.53% and 76.55% respectively, while “hypopneas” exhibit a lower percentage of 70.89%. Based on the appropriate statistical values, we can conclude that for the majority (159 out of 212) of the patients – regardless the SAS severity of their final diagnosis – more than 80% of apneic episodes are accompanied by an



**Fig. 8** Dataset statistics with respect to the scored ODI per patient. (a) Relationship between the ODI and the AHI for all patients in the dataset. While the ascending behavior of the ODI with the increase of the AHI is obvious, the goodness of fit with a linear model expressed by 4 factors is SSE: 5.881e + 04, R-square: 0.6801, Adjusted R-square: 0.6786, RMSE: 16.73. (b) The distribution of ODI for patients belonging to different SAS severity class.

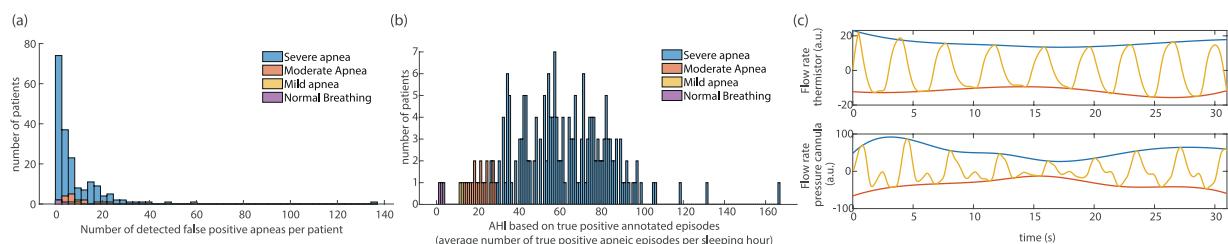


**Fig. 9** Dataset statistics with respect to the annotated relative oxygen desaturation events and arousals. (a) Distribution of time of occurrence of oxygen desaturation events with respect to the associated apnea related episode onset for the main types of apnea/hypopnea (Obstructive, Central, Mixed apnea and Hypopnea). (b) Distribution of the time of arousal occurrence with respect to the end time of the associated apnea, illustrated separately for each type of apnea.

immediate arousal (see Supplementary Fig. 5), while the time of occurrence of the arousal is by average located 2.5 s after the end of the apneic episode. The arousal onset time does not seem to depend on the type of preceding apnea, since the distribution of the time of occurrence for each type separately indicates similar mean values for all types (Fig. 9b).

**Subjectivity of manual labelling of apnea/hypopnea events: How significant is the human induced error in SAS severity estimation.** With the scoring of the apneic events being performed manually by the simultaneous observation of multiple channel, we do expect that a human error is induced in the dataset scoring<sup>98</sup>. This error should be mainly attributed to the inability of the doctor to precisely determine the degree of air flow reduction by visual observation of the flow rate signals (signal label: ‘Flow Patient’ corresponding to a thermistor and a pressure cannula sensor). For the case of a candidate apneic/hypopneic episode presenting a moderate reduction (close to 30%) in the flow rate, the doctors are forced to label the event as positive apneic to eliminate the risk of false negative diagnosis of the SAS severity of the patient. Consequently, an increased number of false positive annotated apneas are anticipated in the dataset.

In order to examine the degree of human induced error, we developed a simple algorithm that automatically checks the compliance of the annotated events’ characteristics with the recommended criteria for apnea scoring<sup>1</sup>. The rule being followed requires, for any respiratory event, a reduction in both flow rate sensors at least equal to 30% of the amplitude in the previous state of normal breathing. It is also mandatory for an annotated apnea event to exhibit a dominant frequency within the range of normal breathing rate in order to be excluded from the positive apneas list. A normal breath lasts for approximately 3–5 s. Thus a range of 0.16–0.4 Hz was selected for the breathing rate rule. The algorithmic process, for flow rate amplitude reduction, extracts the signal envelop through Hilbert transformation, in a way that restricted in time changes – less than 3 s in duration – are ignored. By examining a 5 s time frame prior and after the annotated event, we extracted the maximum flow rate amplitude corresponding to normal breathing. The comparison of this normal breathing amplitude with the minimum amplitude detected within the annotated event determines whether the candidate episode should be accepted as positively annotated or should be rejected as apnea-negative.



**Fig. 10** False positive detections due to manual apnea/hypopnea events scoring. (a) Distribution of the number of detected false positive annotated apneas per patient along with the characterization of SAS diagnosis. (b) Distribution of AHI extracted by the true positive apneas per sleeping hour for all patients along with the originally derived diagnosis for SAS severity. (c) Flow rate for an example of a rejected annotated apnea. Reduction in the flow rate within the annotated episode is insufficient to categorize this episode in one of the types of apnea.

The aforementioned process resulted in the rejection of 1711 false positive annotated apneas among the total sum of 49497 events (3.46%). It is important to note that the number of falsely positive annotated apneas does not seem to correlate with the overall diagnosis of the patient (Fig. 10a). The impact of false positive apneas on the SAS severity estimation is minor, with only 4.24% (9 out of the total sum 212) of the patients receiving an overestimated classification of their state of apnea severity. The distribution of AHI extracted by the true positive annotated apnea/hypopnea events along with the delivered to patient SAS severity diagnosis is illustrated in Fig. 10b.

Among the identified false positive apneas the majority belongs to the type of “hypopnea” (1448 events), 257 were classified to “obstructive” apnea, only 5 events were of type “mixed” apnea and 1 event of type “central”. This particularly illustrates the shortcoming in the quantification of the reduction in the airflow by simple observation of the signals by the doctor.

Two major issues concerning interpretation of flow rate signals have been extensively discussed in the literature. In particular, it is reported that the use of only the nasal pressure sensor could lead to an overestimation of apnea events since the air flow through mouth is entirely ignored<sup>99</sup>. It is also reported that pressure sensors are non-linear therefore their use frequently leads to an overestimation of the flow signal amplitude and an erroneously overdetection of hypopneic events<sup>99,100</sup>. The second important aspect affecting interpretation of airflow-related PSG channels is the accuracy limitations of thermistors. They are reported to underestimate flow reduction, through specific flow pattern studies simulating breathing process<sup>101</sup>. This non-linearity of thermal sensors may lead to a severe under-detection of apneic or hypopneic events<sup>90,92</sup>. Though the aforementioned issues still remain under discussion in the literature, the criteria set in this work, for false positive detection counting, concerned both channels with the requirement to exhibit simultaneously a noticeable reduction of amplitude. In the case where only one of the employed sensors monitors a flow reduction, the event was accounted for false positive detection. An exception in this process appears when the signal exhibits very low amplitude, considered as noise. In that case, the corresponding channel was entirely excluded from the investigation, since the sensor could be temporarily dislocated and the corresponding results could be misleading.

While the rejection of the detected false positive annotated apneas slightly alter the final diagnosis for each patient, the extraction of a list of true positive apneas may be valuable for the development of highly accurate automatic systems for SAS estimation. Thus, we opted for the inclusion of additional events annotation (.rml) files in the dataset named with the prefix “clean” for use in comparative studies and particularly in studies aiming at the development of automatic SAS estimation systems based on breathing sound.

## Code availability

In this study we used Sleepware G3 software for all PSG data acquisition and events annotation. The software is provided by Philips Inc. The custom code used in this work refers to: (a) the algorithm for the synchronization of audio recordings with the PSG signals, so that the information of all annotated episodes and sleep stages can be accurately transferred to the audio signals as well and (b) the detection of false positive apnea/hypopnea events due to the inability of the doctors to precisely quantify the reduction in the airflow amplitude measured through the flow rate channels of the PSG (thermistor and pressure cannula sensors). All custom code developed for this study is available online (“code.rar” file) along with the files of the dataset<sup>79</sup>.

Received: 14 December 2020; Accepted: 17 June 2021;

Published online: 03 August 2021

## References

- Berry, R. B. *et al.* The AASM manual for the scoring of sleep and associated events. *Am. Acad. Sleep Med.* **53**, 1689–1699 (2013).
- Tsara, V., Amfilochiou, A., Papagrigorakis, M. J., Georgopoulos, D. & Liolios, E. Guidelines for diagnosis and treatment of sleep-related breathing disorders in adults and children: Definition and classification of sleep related breathing disorders in adults. Different types and indications for sleep studies (Part 1). *Hippokratia* **13**, 187–191 (2009).
- Guilleminault, C. Obstructive sleep apnea. *The clinical syndrome and historical perspective*. *Med. Clin. North Am.* **69**, 1187–1203, [https://doi.org/10.1016/S0025-7125\(16\)30982-8](https://doi.org/10.1016/S0025-7125(16)30982-8) (1985).
- Young, T., Evans, L., Finn, L. & Palta, M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep* **20**, 705–706, <https://doi.org/10.1093/sleep/20.9.705> (1997).
- Brunetti, L. *et al.* Prevalence of obstructive sleep apnea syndrome in a cohort of 1,207 children of Southern Italy. *Chest* **120**, 1930–1935, <https://doi.org/10.1378/chest.120.6.1930> (2001).

6. Bixler, E. O., Vgontzas, A. N., Have, T. T., Tyson, K. & Kales, A. Effects of age on sleep apnea in men. *Pneumologie* **52**, 467–468, <https://doi.org/10.1164/ajrccm.157.1.9706079> (1998).
7. Resta, O. *et al.* Gender, age and menopause effects on the prevalence and the characteristics of obstructive sleep apnea in obesity. *Eur. J. Clin. Invest.* **33**, 1084–1089, <https://doi.org/10.1111/j.1365-2362.2003.01278.x> (2003).
8. Anuntaseree, W., Rookkapan, K., Kuasirikul, S. & Thongsuksai, P. Snoring and obstructive sleep apnea in Thai school-age children: Prevalence and predisposing factors. *Pediatr. Pulmonol.* **32**, 222–227, <https://doi.org/10.1002/ppul.1112> (2001).
9. Larsson, L. G., Lindberg, A., Franklin, K. A. & Lundbäck, B. Gender differences in symptoms related to sleep apnea in a general population and in relation to referral to sleep clinic. *Chest* **124**, 204–211, <https://doi.org/10.1378/chest.124.1.204> (2003).
10. Bonsignore, M. R., Saaresranta, T., Riha, R. L., Riha, R. & Bonsignore, M. Sex differences in obstructive sleep apnoea. *Eur. Respir. Rev.* **28**, 1–11, <https://doi.org/10.1183/16000617.0030-2019> (2019).
11. Appleton, S. *et al.* Influence of gender on associations of obstructive sleep apnea symptoms with chronic conditions and quality of life. *Int. J. Environ. Res. Public Health* **15**, <https://doi.org/10.3390/ijerph15050930> (2018).
12. Gislason, T., Almqvist, M., Eriksson, G., Taube, A. & Boman, G. Prevalence of sleep apnea syndrome among Swedish men—an epidemiological study. *J. Clin. Epidemiol.* **41**, 571–576, [https://doi.org/10.1016/0895-4356\(88\)90061-3](https://doi.org/10.1016/0895-4356(88)90061-3) (1988).
13. Lopez, P. P., Stefan, B., Schulman, C. I. & Byers, P. M. Prevalence of sleep apnea in morbidly obese patients who presented for weight loss surgery evaluation: more evidence for routine screening for obstructive sleep apnea before weight loss surgery. *Am. Surg.* **74**, 834–838 (2008).
14. Romero-Corral, A., Caples, S. M., Lopez-Jimenez, F. & Somers, V. K. Interactions between obesity and obstructive sleep apnea. *Chest* **137**, 711–719, <https://doi.org/10.1378/chest.09-0360> (2010).
15. Jehan, S. *et al.* Obstructive sleep apnea and obesity: implications for public health. *Sleep Med. Disord. Int. J.* **1**, 1–15 (2017).
16. Wolk, R., Shamsuzzaman, A. S. M. & Somers, V. K. Obesity, sleep apnea, and hypertension. *Hypertension* **42**, 1067–1074, <https://doi.org/10.1161/01.HYP.0000101686.98973.A3> (2003).
17. Valencia-Flores, M. *et al.* Prevalence of sleep apnea and electrocardiographic disturbances in morbidly obese patients. *Obes. Res.* **8**, 262–269, <https://doi.org/10.1038/oby.2000.31> (2000).
18. Fletcher, E. C., DeBehnke, R. D., Lovoi, M. S. & Gorin, A. B. Undiagnosed sleep apnea in patients with essential hypertension. *Ann. Intern. Med.* **103**, 190–195, <https://doi.org/10.7326/0003-4819-103-2-190> (1985).
19. Fletcher, E. C. The relationship between systemic hypertension and obstructive sleep apnea: Facts and theory. *Am. J. Med.* **98**, 118–128, [https://doi.org/10.1016/S0002-9343\(99\)80395-7](https://doi.org/10.1016/S0002-9343(99)80395-7) (1995).
20. Hla, K. M. *et al.* Sleep apnea and hypertension: a population based study. *Ann Intern Med* **120**, 382–388, <https://doi.org/10.7326/0003-4819-120-5-199403010-00005> (1994).
21. Worsnop, C. J. *et al.* The prevalence of obstructive sleep apnea in hypertensives. *Pneumologie* **52**, 469, <https://doi.org/10.1164/ajrccm.157.1.9609063> (1998).
22. Lam, D. C. L. *et al.* Prevalence and recognition of obstructive sleep apnea in Chinese patients with type 2 diabetes mellitus. *Chest* **138**, 1101–1107, <https://doi.org/10.1378/chest.10-0596> (2010).
23. Reichmuth, K. J., Austin, D., Skatrud, J. B. & Young, T. Association of sleep apnea and type II diabetes: A population-based study. *Am. J. Respir. Crit. Care Med.* **172**, 1590–1595, <https://doi.org/10.1164/rccm.200504-637OC> (2005).
24. Kent, B. D. *et al.* Diabetes mellitus prevalence and control in sleep-disordered breathing: The European Sleep Apnea Cohort (ESADA) study. *Chest* **146**, 982–990, <https://doi.org/10.1378/chest.13-2403> (2014).
25. Einhorn, D. *et al.* Prevalence of sleep apnea in a population of adults with type 2 diabetes mellitus. *Endocr. Pract.* **13**, 355–362, <https://doi.org/10.4158/EP13.4.355> (2007).
26. Dyken, M. E. & Im, K. B. Obstructive sleep apnea and stroke. *Chest* **136**, 1668–1677, <https://doi.org/10.1378/chest.08-1512> (2009).
27. Johnson, K. G. & Johnson, D. C. Frequency of sleep apnea in stroke and TIA patients: A meta-analysis. *J. Clin. Sleep Med.* **6**, 131–137, <https://doi.org/10.5664/jcsm.27760> (2010).
28. Dziewras, R. *et al.* Increased prevalence of sleep apnea in patients with recurring ischemic stroke compared with first stroke victims. *J. Neurol.* **252**, 1394–1398, <https://doi.org/10.1007/s00415-005-0888-7> (2005).
29. Tosun, A., Köktürk, O., Karatas, G. K., Çiftçi, T. U. & Sepici, V. Obstructive sleep apnea in ischemic stroke patients. *Clinics* **63**, 625–630, <https://doi.org/10.1590/s1807-59322008000500010> (2008).
30. Butt, M., Dwivedi, G., Khair, O. & Lip, G. Y. H. Obstructive sleep apnea and cardiovascular disease. *Int. J. Cardiol.* **139**, 7–16, <https://doi.org/10.1016/j.ijcard.2009.05.021> (2010).
31. Bauters, F., Rietzschel, E. R., Hertegonne, K., Borch, C. & Chirinos, J. A. The link between obstructive sleep apnea and cardiovascular disease. *Curr. Atheroscler. Rep.* **18**, 1–11, <https://doi.org/10.1007/s11883-015-0556-z> (2016).
32. Lanfranchi, P. A. *et al.* Central sleep apnea in left ventricular dysfunction: Prevalence and implications for arrhythmic risk. *Circulation* **107**, 727–732, <https://doi.org/10.1161/01.cir.0000049641.11675.ee> (2003).
33. Vozoris, N. T. Sleep apnea-plus: Prevalence, risk factors, and association with cardiovascular diseases using United States population-level data. *Sleep Med.* **13**, 637–644, <https://doi.org/10.1016/j.sleep.2012.01.004> (2012).
34. Kato, M., Adachi, T., Koshibo, Y. & Somers, V. K. Obstructive sleep apnea and cardiovascular disease. *Circ. J.* **73**, 1363–1370, <https://doi.org/10.1253/circj.cj-09-0364> (2009).
35. Martínez-García, M. Á., Campos-Rodríguez, F. & Farré, R. Sleep apnoea and cancer: Current insights and future perspectives. *Eur. Respir. J.* **40**, 1315–1317, <https://doi.org/10.1183/09031936.00127912> (2012).
36. Engleman, H. M. & Douglas, N. J. Sleep · 4: Sleepiness, cognitive function, and quality of life in obstructive apnoea/hypopnoea syndrome. *Thorax* **59**, 618–622, <https://doi.org/10.1136/thx.2003.015867> (2004).
37. Engleman, H. & Joffe, D. Neuropsychological function in obstructive sleep apnoea. *Sleep Med. Rev.* **3**, 59–78, [https://doi.org/10.1016/s1087-0792\(99\)90014-x](https://doi.org/10.1016/s1087-0792(99)90014-x) (1999).
38. Lacasse, Y., Godbout, C. & Sériès, F. Health-related quality of life in obstructive sleep apnoea. *Eur. Respir. J.* **19**, 499–503, <https://doi.org/10.1183/09031936.002.00216902> (2002).
39. Horstmann, S., Hess, C. W., Bassetti, C., Gugger, M. & Mathis, J. Sleepiness-related accidents in sleep apnea patients. *Sleep* **23**, 1–7 (2000).
40. Haraldsson, P.-O., Carefolt, C., Diderichsen, F., Nygren, A. & Tingvall, C. Clinical symptoms of sleep apnea syndrome and automobile accidents. *ORL* **52**, 57–62, <https://doi.org/10.1159/000276104> (1990).
41. Pack, A. I., Dinges, D. & Maislin, G. A study of prevalence of sleep apnea among commercial truck drivers. Report No. FMCSA-RT-02-030 (*American Trucking Associations Foundation*, 2001).
42. Sia, C. H. *et al.* Awareness and knowledge of obstructive sleep apnea among the general population. *Sleep Med.* **36**, 10–17, <https://doi.org/10.1016/j.sleep.2017.03.030> (2017).
43. Punjabi, N. M. The epidemiology of adult obstructive sleep apnea. *Proc. Am. Thorac. Soc.* **5**, 136–143, <https://doi.org/10.1513/pats.200709-155MG> (2008).
44. Peppard, P. E. *et al.* Increased prevalence of sleep-disordered breathing in adults. *Am. J. Epidemiol.* **177**, 1006–1014, <https://doi.org/10.1093/aje/kws342> (2013).
45. Benjafield, A. V. *et al.* Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir. Med.* **7**, 687–698, [https://doi.org/10.1016/S2213-2600\(19\)30198-5](https://doi.org/10.1016/S2213-2600(19)30198-5) (2019).
46. Senaratna, C. V. *et al.* Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Med. Rev.* **34**, 70–81, <https://doi.org/10.1016/j.smrv.2016.07.002> (2017).
47. Ho, M. L. & Brass, S. D. Obstructive sleep apnea. *Neurol. Int.* **3**, 60–67, <https://doi.org/10.4081/ni.2011.e15> (2011).

48. Jafari, B. & Mohsenin, V. Polysomnography. *Clin. Chest Med.* **31**, 287–297, <https://doi.org/10.1016/j.ccm.2010.02.005> (2010).
49. Nandakumar, R., Gollakota, S. & Watson, N. Contactless sleep apnea detection on smartphones. In *MobiSys 2015 - Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 45–57, <https://doi.org/10.1145/2742647.2742674> (2015).
50. Al-Mardini, M., Aloul, F., Sagahyron, A. & Al-Husseini, L. Classifying obstructive sleep apnea using smartphones. *J. Biomed. Inform.* **52**, 251–259, <https://doi.org/10.1016/j.jbi.2014.07.004> (2014).
51. Tanigawa, T. *et al.* Monitoring Sound To Quantify Snoring and Sleep Apnea. *J. Clin. Sleep Med.* **10**, 73–78, <https://doi.org/10.5664/jcsm.3364> (2014).
52. Al-Mardini, M., Aloul, F., Sagahyron, A. & Al-Husseini, L. On the use of smartphones for detecting obstructive sleep apnea. In *13th IEEE International Conference on BioInformatics and BioEngineering, IEEE BIBE 2013*, 13–16, <https://doi.org/10.1109/BIBE.2013.6701674> (2013).
53. Penzel, T., Schöbel, C. & Fietze, I. New technology to assess sleep apnea: Wearables, smartphones, and accessories. *F1000Research* **7**, 1–12, <https://doi.org/10.12688/f1000research.13010.1> (2018).
54. Kaguara, A., Myoung Nam, K. & Reddy, S. A deep neural network classifier for diagnosing sleep apnea from ECG data on smartphones and small embedded systems. *Thesis Swarthmore College* (2015).
55. Tseng, M. H. *et al.* Development of an intelligent app for obstructive sleep apnea prediction on android smartphone using data mining approach. In *Proceedings - IEEE 9th International Conference on Ubiquitous Intelligence and Computing and IEEE 9th International Conference on Autonomic and Trusted Computing, UIC-ATC 2012*, 774–779, <https://doi.org/10.1109/UIC-ATC.2012.89> (2012).
56. Yadollahi, A. & Moussavi, Z. Apnea detection by acoustical means. In *Proceedings of the 28th IEEE EMBS Annual International Conference*, 4623–4626, <https://doi.org/10.1109/EMBS.2006.260391> (2006).
57. Moussavi, Z., Yadollahi, A. & Camorlinga, S. Breathing sound analysis for detection of sleep apnea/hypopnea events. vol. US 7,559,9 (2009).
58. Emoto, T., Abeyratne, U. R., Akutagawa, M., Nagashino, H. & Kinouchi, Y. Feature extraction for snore sound via neural network processing. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 5477–5480, <https://doi.org/10.1109/EMBS.2007.4353585> (2007).
59. Meskanen, M. Apnea detection using a tracheal microphone and a back propagation neural network. *Med. Biol. Eng. Comput.* **34**, 115–116 (1996).
60. Nakano, H., Furukawa, T. & Tanigawa, T. Tracheal sound analysis using a deep neural network to detect sleep apnea. *J. Clin. Sleep Med.* **15**, 1125–1133, <https://doi.org/10.5664/jcsm.7804> (2019).
61. Emoto, T. *et al.* Artificial neural networks for breathing and snoring episode detection in sleep sounds. *Physiol. Meas.* **33**, 1675–1689, <https://doi.org/10.1088/0967-3334/33/10/1675> (2012).
62. Kang, B., Dang, X. & Wei, R. Snoring and apnea detection based on hybrid neural networks. In *Proceedings of the 2017 International Conference on Orange Technologies, ICOT 2017* vols 2018-Janua, 57–60, <https://doi.org/10.1109/ICOT.2017.8336088> (2018).
63. Ichimaru, Y. & Moody, G. B. Development of the polysomnographic database on CD-ROM. *Psychiatry Clin. Neurosci.* **53**, 175–177, <https://doi.org/10.1046/j.1440-1819.1999.00527.x> (1999).
64. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220, <https://doi.org/10.1161/01.cir.101.23.e215> (2000).
65. Zhang, G. Q. *et al.* The National Sleep Research Resource: Towards a sleep data commons. *J. Am. Med. Informatics Assoc.* **25**, 1351–1358, <https://doi.org/10.1093/jamia/ocv064> (2018).
66. Penzel, T., Rg, G. B. M., Goldbergs, M. A. L. & Peter, H. The apnea-ECG database. *Comput. Cardiol.* **27**, 255–258, <https://doi.org/10.1109/CIC.2000.898505> (2000).
67. Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A. C. & Oberényé, J. J. L. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* **47**, 1185–1194, <https://doi.org/10.1109/10.867928> (2000).
68. Quan, S. F. *et al.* The Sleep Heart Health Study: Design, rationale, and methods. *Sleep* **20**, 1077–1085, <https://doi.org/10.1093/sleep/20.12.1077> (1997).
69. Young, T. *et al.* Burden of Sleep Apnea: Rationale, Design, and Major Findings of the Wisconsin Sleep Cohort Study. *WMJ.* **108**, 246–249 (2009).
70. Lee, H. *et al.* NCH Sleep DataBank: a large collection of real-world pediatric sleep studies. arXiv 1–19 Preprint at <https://arxiv.org/abs/2102.13284> (2021).
71. Rosen, C. L. *et al.* Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: Association with race and prematurity. *J Pediatr* **142**, 383–389, <https://doi.org/10.1067/mpd.2003.28> (2003).
72. Facco, F. L. *et al.* NuMoM2b sleep disordered breathing study: objectives and methods. *Am J Obs. Gynecol.* **212**, 542.e1–542.e127, <https://doi.org/10.1016/j.ajog.2015.01.021> (2015).
73. Blackwell, T. *et al.* Associations of sleep architecture and sleep disordered breathing with cognition in older community-dwelling men: the MrOS sleep study. *J Am Geriatr Soc.* **59**, 2217–2225, <https://doi.org/10.1111/j.1532-5415.2011.03731.x> (2011).
74. Foley, D. J. *et al.* Sleep-disordered breathing and cognitive impairment in elderly Japanese-American men. *Sleep* **26**, 596–599, <https://doi.org/10.1093/sleep/26.5.596> (2003).
75. Redline, S. *et al.* The familial aggregation of obstructive sleep apnea. *Am. J. Respir. Crit. Care Med.* **151**, 682–687, [https://doi.org/10.1164/ajrccm/151.3\\_Pt\\_1.682](https://doi.org/10.1164/ajrccm/151.3_Pt_1.682) (1995).
76. Azarbarzin, A. & Moussavi, Z. Snoring sounds variability as a signature of obstructive sleep apnea. *Med. Eng. Phys.* **35**, 479–485, <https://doi.org/10.1016/j.medengphy.2012.06.013> (2013).
77. Janott, C. *et al.* Snoring classified: The Munich-Passau Snore Sound Corpus. *Comput. Biol. Med.* **94**, 106–118, <https://doi.org/10.1016/j.combiomed.2018.01.007> (2018).
78. Castillo-Escario, Y., Ferrer-Lluis, I., Montserrat, J. M. & Jane, R. Automatic silence events detector from smartphone audio signals: a pilot mHealth system for sleep apnea monitoring at home. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* vol. 2019, 4982–4985, <https://doi.org/10.1109/EMBC.2019.8857906> (2019).
79. Korompili, G. *et al.* PSG-Audio (V2). Sci. DataBase <https://doi.org/10.11922/sciencedb.00345> (2020).
80. Lado, M. J. *et al.* Detecting sleep apnea by heart rate variability analysis: Assessing the validity of databases and algorithms. *J. Med. Syst.* **35**, 473–481, <https://doi.org/10.1007/s10916-009-9383-5> (2011).
81. The European Commission. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). vol. 119 (2016).
82. Ciftci, B., Ciftci, T. U. & Guven, S. F. Split-night versus full-night polysomnography: comparison of the first and second parts of the night. *Arch. Bronconeumol. English Ed.* **44**, 3–7, [https://doi.org/10.1016/s1579-2129\(08\)60002-6](https://doi.org/10.1016/s1579-2129(08)60002-6) (2008).
83. Malhotra, R. K. & Avidan, A. Y. S Stages and Scoring Technique. In *Atlas of Sleep Medicine*, 77–99, <https://doi.org/10.1016/B978-1-4557-1267-0.00003-5> (Elsevier Inc., 2014).
84. Kemp, B., Värri, A., Rosa, A. C., Nielsen, K. D. & Gade, J. A simple format for exchange of digitized polygraphic recordings. *Electroencephalogr. Clin. Neurophysiol.* **82**, 391–393, [https://doi.org/10.1016/0013-4694\(92\)90009-7](https://doi.org/10.1016/0013-4694(92)90009-7) (1992).
85. Iber, C. Are we ready to define central hypopneas? *Sleep* **36**, 305–306 (2013).
86. Shamim-Uzzaman, Q. A., Singh, S. & Chowdhuri, S. Hypopnea definitions, determinants and dilemmas: a focused review. *Sleep Sci. Pract.* **2**, 1–12, <https://doi.org/10.1186/s41606-018-0023-1> (2018).

87. Guilleminault, C., Tilkian, A. & Dement, W. C. The sleep apnea syndromes. *Annu. Rev. Med.* **27**, 465–484, <https://doi.org/10.1146/annurev.me.27.020176.002341> (1976).
88. Borsini, E., Nogueira, F. & Nigro, C. Apnea-hypopnea index in sleep studies and the risk of over-simplification. *Sleep Sci.* **11**, 45–48, <https://doi.org/10.5935/1984-0063.20180010> (2018).
89. Punjabi, N. M. Counterpoint: Is the Apnea-Hypopnea Index the best way to quantify the severity of sleep-disordered breathing? *No. Chest* **149**, 16–19, <https://doi.org/10.1378/chest.14-2261> (2016).
90. Nikkinen, S. *et al.* Intra-night variation in apnea-hypopnea index affects diagnostics and prognostics of obstructive sleep apnea. *Sleep Breath.* **24**, 379–386, <https://doi.org/10.1007/s11325-019-01885-5> (2020).
91. Tilkian, A. G. *et al.* Hemodynamics in sleep induced apnea. *Studies during wakefulness and sleep. Ann. Intern. Med.* **85**, 714–719, <https://doi.org/10.7326/0003-4819-85-6-714> (1976).
92. Remmers, J. E., DeGroot, W. J., Sauerland, E. K. & Anch, A. M. Pathogenesis of upper airway occlusion during sleep. *J. Appl. Physiol. Respir. Environ. Exerc. Physiol.* **44**, 931–938, <https://doi.org/10.1152/jappl.1978.44.6.931> (1978).
93. Temirbekoy, D., Gunes, S., Yazici, Z. M. & Sayin, I. The ignored parameter in the diagnosis of obstructive sleep apnea syndrome the Oxygen Desaturation Index. *Turk Otolaryngoloji Arsivi/Turkish Arch. Otolaryngol.* 1–6 (2018).
94. Dos Santos, C., Samuels, M., Laverty, A. & Raywood, E. Comparison of oxygen desaturation index and apnoea-hypopnoea index for categorising OSA in children. *ERS International Congress* vol. 52, PA549, <https://doi.org/10.1183/13993003.congress-2018.PA549> (2018).
95. Nikkinen, S., Afara, I. O., Leppänen, T. & Töyräs, J. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Sci. Rep.* **9**, 1–9, <https://doi.org/10.1038/s41598-019-49330-7> (2019).
96. Coronel, C. *et al.* Detection of respiratory events by respiratory effort and oxygen desaturation. *J. Med. Biol. Eng.* **40**, 517–525, <https://doi.org/10.1007/s40846-020-00524-9> (2020).
97. Kulkas, A., Tiilohonen, P., Julkunen, P., Mervaala, E. & Töyräs, J. Desaturation delay, parameter for evaluating severity of sleep disordered breathing. In *Long M. (eds) World Congress on Medical Physics and Biomedical Engineering May 26–31, 2012, Beijing, China. IFMBE Proceedings* vol. 39, 336– (2010).
98. Collop, N. A. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep Med.* **3**, 43–47, [https://doi.org/10.1016/s1389-9457\(01\)00115-0](https://doi.org/10.1016/s1389-9457(01)00115-0) (2002).
99. Berry, R. B. *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.* **8**, 597–619 (2012).
100. Farré, R., Rigau, J., Montserrat, J. M., Ballester, E. & Navajas, D. Relevance of linearizing nasal prongs for assessing hypopneas and flow limitation during sleep. *Am. J. Respir. Crit. Care Med.* **163**, 494–497, <https://doi.org/10.1164/ajrccm.163.2.2006058> (2001).
101. Farré, R., Montserrat, J. M., Rotger, M., Ballester, E. & Navajas, D. Accuracy of thermistors and thermocouples as flow-measuring devices for detecting hypopneas. *Eur. Respir. J.* **11**, 179–182, <https://doi.org/10.1183/09031936.98.11010179> (1998).

## Acknowledgements

We acknowledge the contribution of Maria Mpregkou, certified technician of the clinical team of Sleep Study Unit of Sismanoglio, in the first step of sleep stages and respiratory events annotation of the PSG studies. This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK- 03957\_ Automatic Pre-Hospital, In-Home, Sleep Apnea Examination).

## Author contributions

Conceptualization G.K., S.A.M., S.M.P.; Experiment design and setup S.A.M., N.A.T., S.M.P.; Medical data collection and interpretation A.A., E.K., C.M.; dataset development and validation G.K., L.K.; Literature review, database maintenance M.K.; Dataset statistical analysis and interpretation G.K.; Article drafting G.K.; Critical revision of the article: A.A., S.M.P.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-00977-w>.

**Correspondence** and requests for materials should be addressed to S.M.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021