



# Screening for obstructive sleep apnea with novel hybrid acoustic smartphone app technology

Roxana Tiron<sup>1#</sup>, Graeme Lyon<sup>1#</sup>, Hannah Kilroy<sup>1</sup>, Ahmed Osman<sup>1</sup>, Nicola Kelly<sup>1</sup>, Niall O'Mahony<sup>1</sup>, Cesar Lopes<sup>1</sup>, Sam Coffey<sup>1</sup>, Stephen McMahon<sup>1</sup>, Michael Wren<sup>1</sup>, Kieran Conway<sup>1</sup>, Niall Fox<sup>1</sup>, John Costello<sup>1</sup>, Redmond Shouldice<sup>1^</sup>, Katharina Lederer<sup>2</sup>, Ingo Fietze<sup>2</sup>, Thomas Penzel<sup>2</sup>

<sup>1</sup>ResMed Sensor Technologies, NexusUCD, Dublin, Ireland; <sup>2</sup>Advanced Sleep Research, Berlin, Germany

*Contributions:* (I) Conception and design: R Tiron, G Lyon, H Kilroy, R Shouldice, M Wren, S McMahon, K Conway, N Fox, J Costello; (II) Administrative support: K Conway, K Lederer; (III) Provision of study materials or patients: N Kelly, S Coffey, K Lederer, I Fietze, T Penzel; (IV) Collection and assembly of data: A Osman, H Kilroy, N O'Mahony, C Lopes; (V) Data analysis and interpretation: G Lyon, R Tiron, H Kilroy, R Shouldice; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

*Correspondence to:* Redmond Shouldice, PhD, MBA. ResMed Sensor Technologies, NexusUCD, Block 9 & 10, Belfield Office Park, Dublin 4, Ireland. Email: redmond.shouldice@resmed.com.

**Background:** Obstructive sleep apnea (OSA) has a high prevalence, with an estimated 425 million adults with apnea hypopnea index (AHI) of  $\geq 15$  events/hour, and is significantly underdiagnosed. This presents a significant pain point for both the sufferers, and for healthcare systems, particularly in a post COVID-19 pandemic world. As such, it presents an opportunity for new technologies that can enable screening in both developing and developed countries. In this work, the performance of a non-contact OSA screener App that can run on both Apple and Android smartphones is presented.

**Methods:** The subtle breathing patterns of a person in bed can be measured via a smartphone using the “Firefly” app technology platform [and underpinning software development kit (SDK)], which utilizes advanced digital signal processing (DSP) technology and artificial intelligence (AI) algorithms to identify detailed sleep stages, respiration rate, snoring, and OSA patterns. The smartphone is simply placed adjacent to the subject, such as on a bedside table, night stand or shelf, during the sleep session. The system was trained on a set of 128 overnights recorded at a sleep laboratory, where volunteers underwent simultaneous full polysomnography (PSG), and “Firefly” smartphone app analysis. A separate independent test set of 120 recordings was collected across a range of Apple iOS and Android smartphones, and withheld for performance evaluation by a different team. An operating point tuned for mid-sensitivity (i.e., balancing sensitivity and specificity) was chosen for the screener.

**Results:** The performance on the test set is comparable to ambulatory OSA screeners, and other smartphone screening apps, with a sensitivity of 88.3% and specificity of 80.0% [with receiver operating characteristic (ROC) area under the curve (AUC) of 0.92], for a clinical threshold for the AHI of  $\geq 15$  events/hour of detected sleep time.

**Conclusions:** The “Firefly” app based sensing technology offers the potential to significantly lower the barrier of entry to OSA screening, as no hardware (other than the user’s personal smartphone) is required. Additionally, multi-night analysis is possible in the home environment, without requiring the wearing of a portable PSG or other home sleep test (HST).

**Keywords:** Sleep-disordered breathing (SDB); obstructive sleep apnea (OSA); screening; smartphone; apnea hypopnea index (AHI)

<sup>^</sup> ORCID: 0000-0002-0516-2266.

Submitted Feb 08, 2020. Accepted for publication Jul 03, 2020.

doi: 10.21037/jtd-20-804

View this article at: <http://dx.doi.org/10.21037/jtd-20-804>

## Introduction

Sleep disordered breathing (SDB) is characterized by abnormal respiratory patterns, which can include pauses in breathing, and insufficient ventilation during sleep. The most common types of SDB include upper airway resistance syndrome (UARS), and obstructive sleep apnea (OSA). OSA can be asymptomatic or symptomatic, accompanied by major neurocognitive and cardiovascular sequelae (i.e., a condition which is the consequence of a previous disease), and typically associated with disrupted sleep, daytime fatigue, and a concomitant reduced quality of life (QOL) (1-3).

Globally, over 936 million adults aged 30–69 years are believed to have mild, moderate or severe OSA [i.e., apnea hypopnea index (AHI) threshold values of  $\geq 5$  events per hour]; 425 million of these adults have either moderate to severe obstructive sleep (AHI  $\geq 15$ ). For example, the number of individuals with AHI  $\geq 15$  is estimated to be 66 million in China, 29 million in India, 25 million in Brazil, 24 million in the USA, and 20 million in Russia (1). Positive airway pressure (PAP) is recommended in patients with excessive sleepiness, impaired sleep related QOL, and comorbid hypertension, which are more likely with an Apnea Hypopnea Index of  $\geq 15$  (1,4).

As the prevalence of OSA rises, so do the associated health, safety, and economic consequences, and the urgent need to identify those suffering with the condition. Access to “gold standard” type I in-lab attended polysomnography (PSG) and associated diagnostic services remains a challenge, particularly in the developing world, and in remote communities (5). As a result, there is growing interest in home monitoring, such as type II portable monitors (PSG), type III & IV portable monitors, screening questionnaires & clinical prediction tools, and well as emerging contactless technologies—particularly for detecting severe OSA in those without severe comorbidity (where lab PSG may still be required) (5,6). In this work, an example of such non-contact monitoring technology is presented.

Existing contact-based monitoring approaches for detecting OSA include a diverse range of sensing and processing; examples of these using sound include use of a microphone attached to the face or to bedding, such as: a

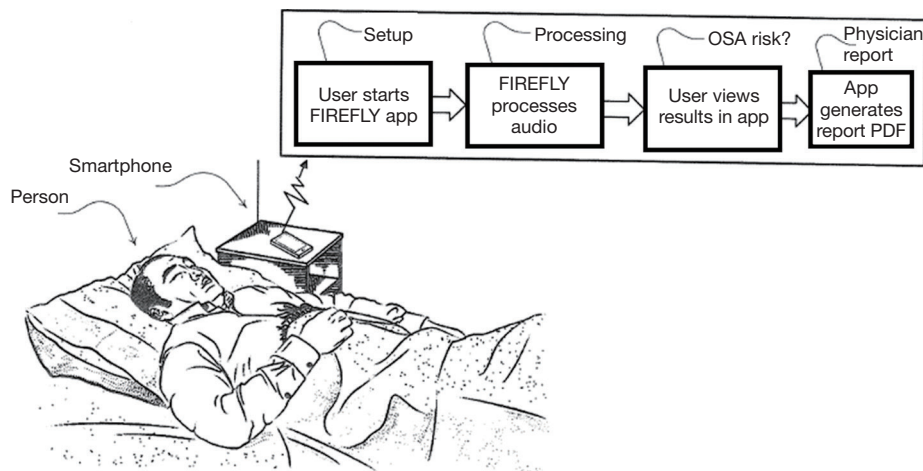
face frame mask (7), a mattress overlay (8), two microphones (one attached to upper lip, the other for ambient noise) (9,10), and oro-nasal based airflow acoustics (10,11).

Wireless (i.e., contactless) recording of sleep and related biometrics has the potential to enable new insights into sleep disorders (12). Fully contactless approaches (i.e., non-contact sensing at a distance, such as where the sensing device is placed on a bedside locker, shelf or similar adjacent to the bed) have been explored to estimate wake & sleep stages (13-16), breathing rate (17-20), and OSA in (21-24), using specialized radio frequency (RF) sensors and artificial intelligence/machine learning (AI/ML) algorithms.

The way people access information relating to their health is rapidly evolving, and increasingly smart devices are playing a bigger role. In 2018, 5.1 billion people (67% of the global population) had subscribed to mobile services, with a further 710 million people expected to subscribe to mobile services for the first time by 2025 (rising to 71% of the population) (25). There were an estimated 3.2 billion smartphone users in 2019 (26)—exceeding 40% of the world population (27). The number of unique mobile subscribers in North America was 300 million in early 2018 (84% of the population) and is growing steadily (28).

This rapid growth has fueled interest in utilizing the smartphone sensors and platforms for new sensing and health analysis. For example, there has been recent work on using passive acoustic analysis for detecting snore and OSA (29,30) and sonar reflections for detecting OSA (31,32), as well as sleep staging (33), using smartphones. However, the former can suffer from interference from background ambient noise, and the latter can require very specific setup to get accurate signals, with limited smartphone support. Additionally, these approaches can estimate sleep quality or OSA risk—but seldom do both simultaneously.

In this work, a novel hybrid machine learning software development kit (SDK) using both active sonar and passive acoustic analysis that can overcome these challenges is evaluated. This approach can estimate an AHI value, and screen for AHI  $\geq 15$  events per hour of sleep time. This is implemented as both an Apple (Cupertino, CA, USA) iOS app (Swift or Objective C) as well as Google (Mountain View, CA, USA) Android (Java) app, wrapping a C++ SDK. This patent pending “Firefly” technology does not require



**Figure 1** High level overview of the Firefly system, showing a smartphone placed beside the bed.

any custom hardware, and is compatible with a wide range of current and future smartphones, with additional potential applications in tablets, smart TVs, smart speakers and autonomous vehicles (34-36).

## Methods

### *System overview and typical use case*

The Firefly mobile app (and underpinning SDK) is designed to estimate AHI, and to screen for OSA risk. The technology operates on a mobile computing device such as an Apple or Android smartphone. It works by estimating sleep and breathing patterns, and then analyzing these results in order to track sleep-related health risks associated with sleep apnea.

The hybrid processing approach utilizes:

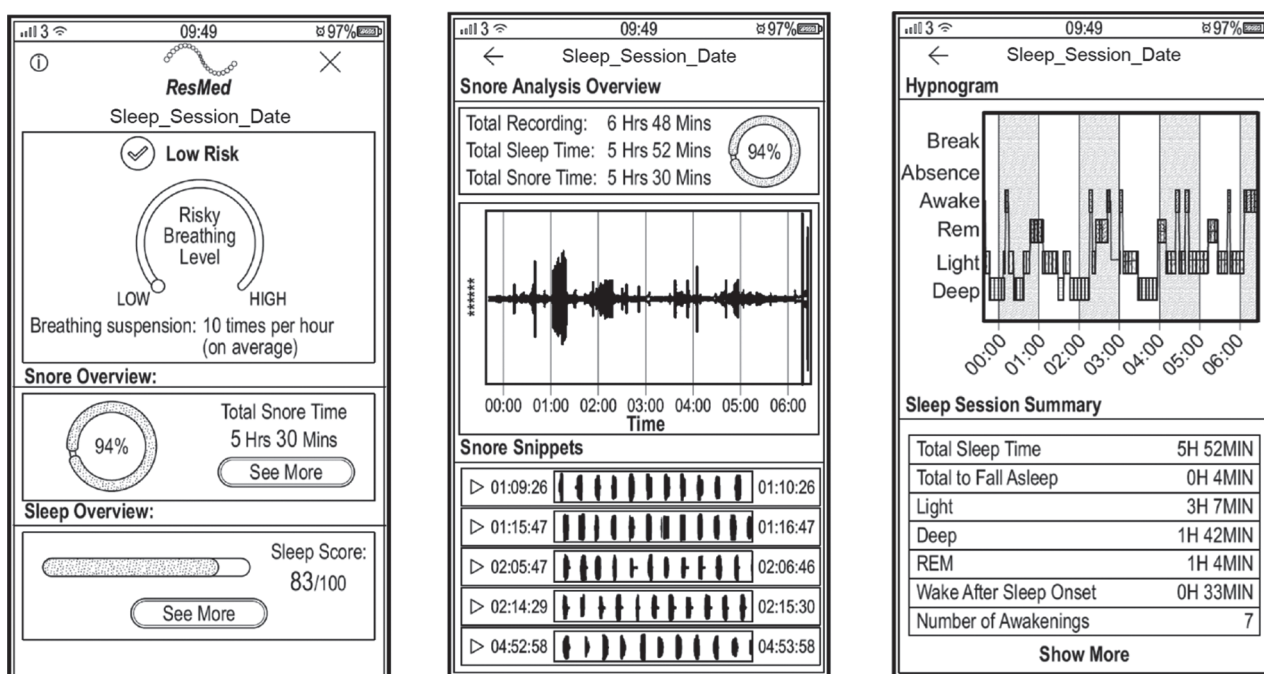
- (I) Passive breath sound detection (and pauses of same, indicative of apnea or hypopnea), and;
- (II) Active “sonar” (comparable to echo location used by bats and dolphins) to detect movement of the person as they move and breathe (and reductions in same, indicative of apnea or hypopnea).

This includes pre-processing, separating, demodulating, and analyzing the complex multipath received signals. Breathing movements during sleep are recorded using the speaker and microphone within the mobile device by detecting changes in the echo of a specially crafted phase aligned modulated signal played via a selected speaker. Breathing sounds during sleep, background ambient sounds, as well as the multipath reflections of the transmitted signal,

are recorded using a selected microphone within the mobile device.

The mobile app development process includes steps of prototyping in MATLAB (Mathworks, Natick, MA), Python & Tensorflow (37), performance evaluation, followed by implementation in C++, equivalence testing, and test platform evaluation on Windows, Linux and MacOS using continuous integration. The resulting apps are native Apple iOS Swift (with portions of Objective C) and Android Java apps, wrapping the compiled C++ components. This approach allows a common Firefly compiled core to be used cross-platform, and facilitates rapid updates to the core sensing and algorithms. The development follows software life cycle processes based on best practice (e.g., IEC 62304), agile software development, and risk management processes (e.g., ISO 14971).

A typical use case is as follows. The user installs the app on their Android or Apple phone from an app store. No other equipment is required, and no phone modification is needed; i.e., the user just needs to have the app installed on a regular smartphone. The phone is placed on a bedside table. It is recommended that the phone is placed above the mattress height, at about an arm’s length away from the person, and with the charger port pointed towards the chest (which is the area where the speaker and mic are located in modern phones). The app is started at the beginning of the night, automatically commences data processing, and is stopped at the end of the night (or stops automatically)—see *Figures 1,2*. If the user picks up or interacts with the phone (based on processing of gyroscope, accelerometer, charging status, and user interface (UI) touches), the sensing is



**Figure 2** Wireframes of possible app screens. These illustrate an estimated AHI of 10 (below a threshold of  $AHI \geq 15$ ), although the subject was snoring for most of the night (94%), and should probably continue to monitor their sleep as a result. AHI, apnea hypopnea index.

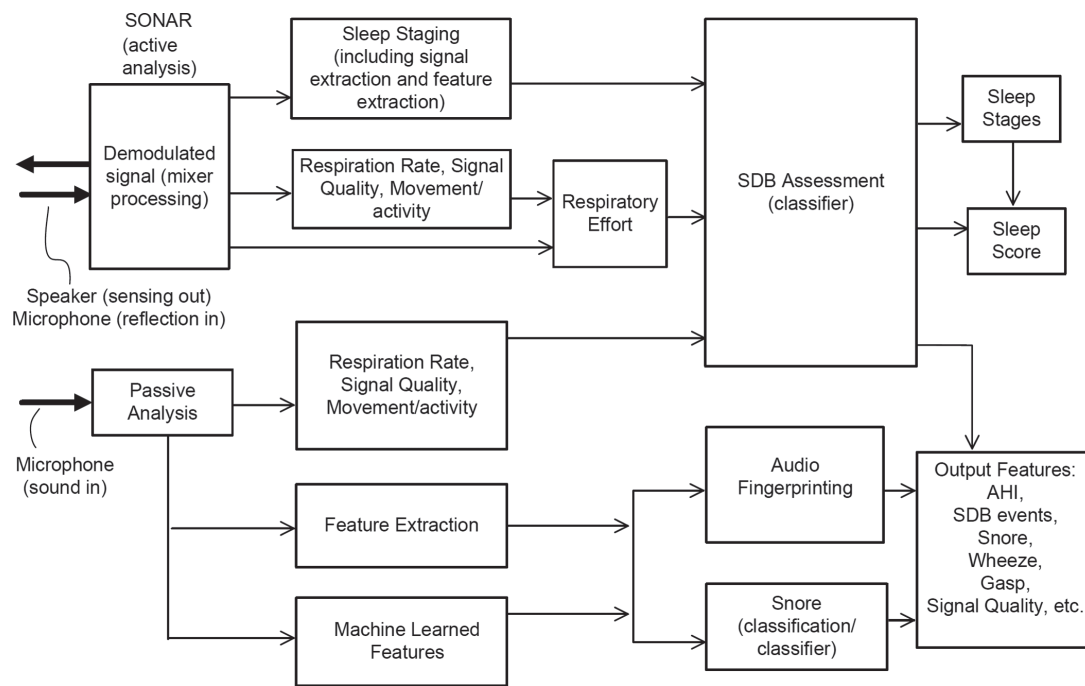
paused for the duration of the interaction. In addition, if the smartphone is connected to an external Bluetooth speaker or ear buds, the sensing is paused.

The algorithm uses audio digital signal processing (DSP) of sound in the room as detected by the phone's microphone sampled at 48 kHz, including breathing, the sound of snoring, and of breathing pauses. This includes analysis of breath cessation (quiet time) or breath reduction between breaths and/or snores, and recovery breath gasp after an apnea. It also employs inaudible sonar (using a range above 18 kHz) to measure the breath parameters (modulated breathing patterns such as clusters of gaps and reductions due to apneas and hypopneas). This hybrid approach increases the robustness of the system, and allows separation of the biometrics of nearest person to the smartphone from a bed partner. In terms of privacy, as the audio data are processed by the production Firefly SDK locally on the smartphone, no such data need be transmitted to any third party system.

During development, many different real-world bedroom scenarios were evaluated and tested, in terms of smartphone type, smartphone placement beside the bed (such as height, distance, and angle), impact of different bedding materials, and background sounds. Different factors such as age, body

mass index (BMI), and gender were checked in terms of movement and respiration pattern, as well as the impact of one or two people in bed on the detected waveforms. Lab based testing was carried out with custom robotic breathing phantoms, in acoustically treated environments, with calibrated equipment. For example, (I) distance impact was evaluated to test respiration rate and movement of subject at various distances and heights to a smartphone device under test (DUT) with a typical effective range of 1 m, (II) angle tests of the respiration signal at various angles to the DUT chest location covered (from 0–360° angle, with optimal results at  $\pm 60^\circ$ ), and (III) blanket impact tests of respiration rate with subject covered by selection of different bed linens (such as no blanket, sheet, thin blanket, quilt, feather quilt, or thick cotton blanket, multiple blankets, and so forth). Older and current generation phones from suppliers such as Apple (e.g. from iPhone 6 to 11 Pro Max), Samsung (S6 and later), Google (Pixel family), Huawei, Motorola, LG and others were evaluated; their speaker and mic sub systems were characterized in order to determine capability, manufacturing variability, appropriate audio configurations, and suitability of active and passive sensing.

The data are processed using proprietary Firefly AI/ML



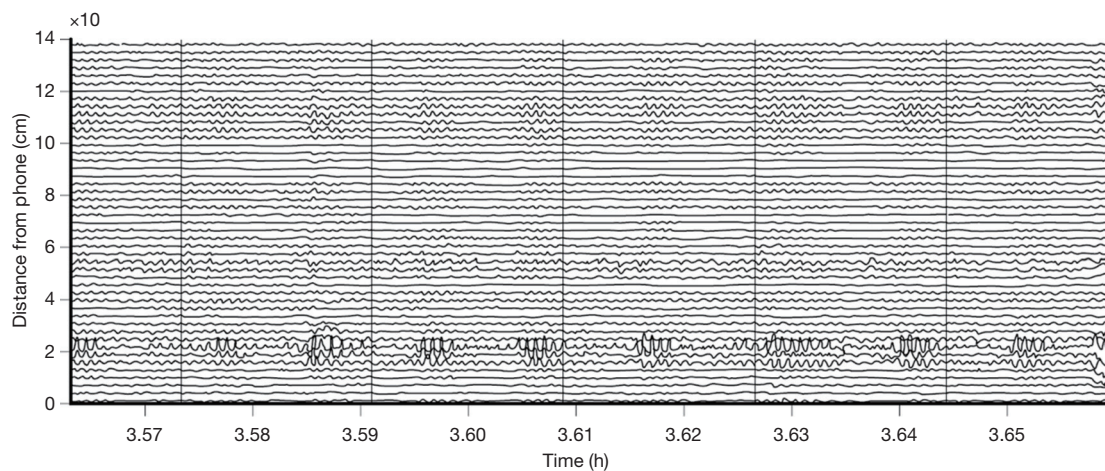
**Figure 3** Firefly app algorithm architecture. Active sonar (using the smartphone speaker to play an inaudible signal, and microphone to receive reflections of that signal) is processed using logistic models, and then fused with simultaneous processing of the microphone full band signal using a neural network to reject non snore sounds.

algorithms, described below and in (32,34,35). An outline of the algorithm architecture is presented in *Figure 3*. Firefly does not change the capabilities of the smartphone speaker; rather, it utilizes the fact that the loudspeaker and microphone components and associated firmware in modern smartphones can support frequencies above those audible to most people. Specifically, it uses the frequency range of 18–20 or 20–22 kHz for sensing using a swept sinusoidal waveform. The reflected signal from the room and the direct path component are synchronized and demodulated. The components of the remaining signal will depend on whether the target is static or moving. Assuming an oscillating target at a given distance (e.g., a potential “breathing” signal), the signal will contain a “beat” signal and a Doppler component. The beat signal is due to the difference in frequency position of the transmit and received sweeps caused by the time delayed reflection from the target. The Doppler is a frequency shift caused by the moving target. The waveform shape and breathing signal is estimated over range, as depicted in *Figure 4*, where the breathing pattern of the person in bed is detected at around 25 cm away from the phone. Firefly is designed to manage the sometimes highly nonlinear lumped response of speaker

and mic at these frequencies; this behavior varies between phone models (and transducer suppliers), and can also vary during a night due to thermal variations, auto gain control (AGC), speaker management, and other sophisticated software control systems running on different phone handsets. By design, the sonar signal is outside the range of speech and typical sounds, but has further trend analysis, de-noising, and smoothing applied to minimize the impact of sharp noises such as pops and bangs in the bedroom environment (e.g., doors closing).

The active sonar signal processing involves further range-bin spectral and morphological processing in order to estimate an effort signal, and then automatically selecting a candidate envelope respiratory effort signal based on signal to noise ratio. Features in the 25–100 sec timescale, related to potential groups of apneas or hypopneas, are provided as inputs to a logistical model. This approach is designed to reject mechanical periodic sources such as fans, and to emphasize SDB patterns.

The passive signal as recorded by the microphone is band-pass filtered in the range 250–8,000 Hz [i.e., to remove the sonar signal, many potentially interfering sounds, and low frequency sounds including mains (AC



**Figure 4** Various motion signals relating to respiratory effort in different range bins are depicted, allowing isolation of one or more people in the active sonar sensing field (shown to 1.4 m from the smartphone in this example, with the person at a distance of 25 cm).

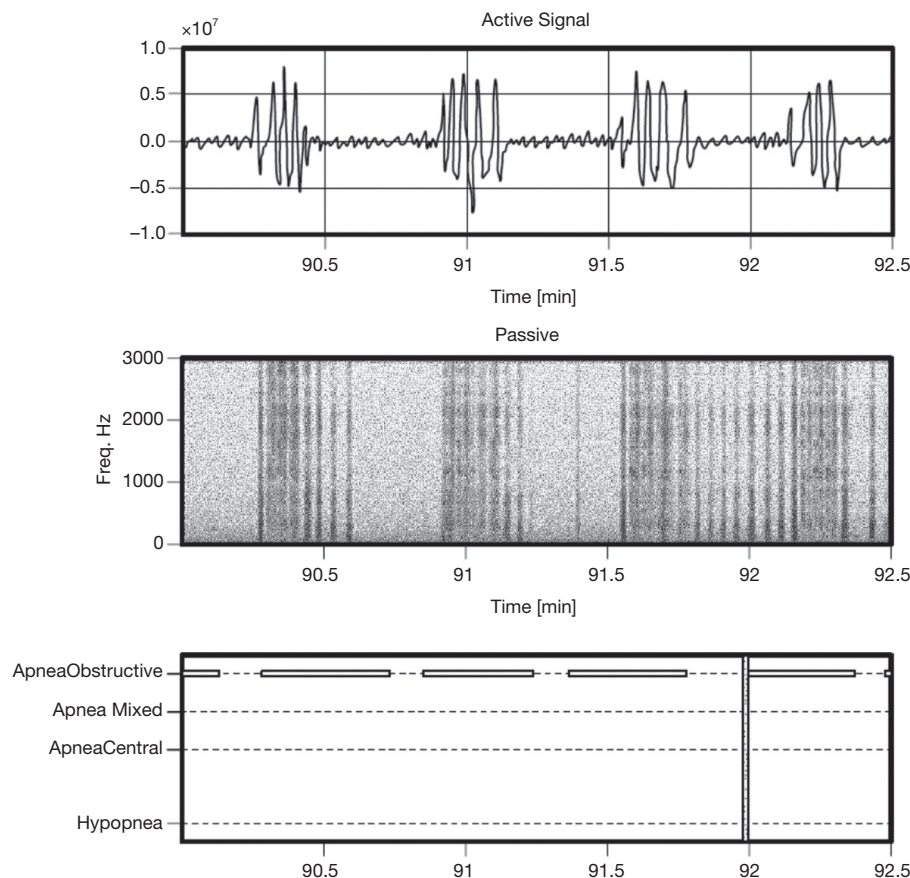
50/60 Hz) hum]. Subsequent steps of envelope processing including peak enhancement, followed by spectral analysis, are performed to produce further features for the logistical models.

The passive signal further includes de-noising, followed by calculating Mel Frequency Cepstrum Coefficients (MFCCs), and then using multiple 2D CNN convolution networks to estimate the probability of particular audio fingerprints such as snore, and the relative intensity of snoring. For snoring, the passive audio is low pass filtered (to 16 kHz), and MFCC's are calculated for every 25 ms of audio. An averaged RMS (root mean square) is also calculated for display purposes, e.g., an approximate representation of intensity of snoring in a graph. Tensorflow is used to train AI models on a large labelled dataset of snore and non-snore audio (including examples of CPAP, ocean, rainfall, city sounds, electro mechanical time switch, large fan, small fan, empty room, home sounds, baby, cot mobile, lawn mower, smoked detector, car alarm, squeaky door, ultrasonic "pest repeller", air conditioning, white noise, pink noise, TV, washing machine, and more). The CNN uses at least two convolution layers (e.g., some standard fully connected ReLU/sigmoid layers), dense layers with softmax classification, and regularization. As these contain more than three layers (including input and output layers), they can be considered deep neural networks, and perform the audio fingerprinting (rejection of background audio) and snore detection.

The combination of the sonar detection of bio-motion (including breathing) and processing of the breathing sound

is used to estimate high and low probability areas of OSA in an overnight recording. Input features are generated based on knowledge that obstructive events typically occur in sequences such as in clusters of modulated breathing, and that there are disruptions or gaps due to the obstructive apnea or hypopnea in the acoustic signal. While Firefly does not measure either airflow or drop in oxygen saturation directly, it does measure breathing motion via the sonar stream, and thus can capture variation in respiratory effort. Since clinical thresholds for variation in respiratory effort that would allow scoring of hypopneas do not exist, the system does not detect hypopneas in an event by event manner. Instead, Firefly captures periods in the sleep session where variability in respiratory effort is detected, consistent with clusters of hypopneas occurring. Analogous cluster behavior has been seen in heart rate and oximetry signals for example (38,39). A hybrid approach is utilized by Firefly whereby time-frequency features characterizing the frequency, duration, and severity of disruptions across the night are extracted for each recording from both the passive audio, and the sonar reflections, weighted based on the channel with higher respiratory signal quality in a rolling window.

Examples of Firefly detecting obstructive and central events, as well as hypopneas, are provided in *Figures 5-8*. Examples of a full night view (and then two 1 hour chunks) of estimated respiration rate, activity, respiratory effort, acoustic level with overlaid SDB sections, wake, and snore are provided in *Figures 9-11*. For AHI detection and audio fingerprinting, demographic variables were not required,



**Figure 5** Estimated active (sonar) and passive (acoustic) signals for a segment of obstructive apnea (PSG annotations in lower panel) for a subject with a very high AHI of 91.4 events per hour. PSG, polysomnography; AHI, apnea hypopnea index.

and are not used in the models. However, sleep staging does use age and gender to allow for age related changes in sleep architecture.

Sleep staging classification is provided by Firefly AI, which outputs 30 second epoch labels of absence, presence, wake, NREM sleep (N1+N2 light sleep, N3 deep sleep), and REM; the details are discussed in (33). The resulting “time asleep” data are used as part of the AHI calculation, i.e., the duration of time with high probability of apnea or hypopneas (modulation) for total sleep time, while the user is present within the field of the Firefly enabled smartphone.

The sleep staging adds an extra layer of robustness to the system, by rejecting “wake” events (other than arousals due to apneas) that might otherwise be incorrectly detected as apnea.

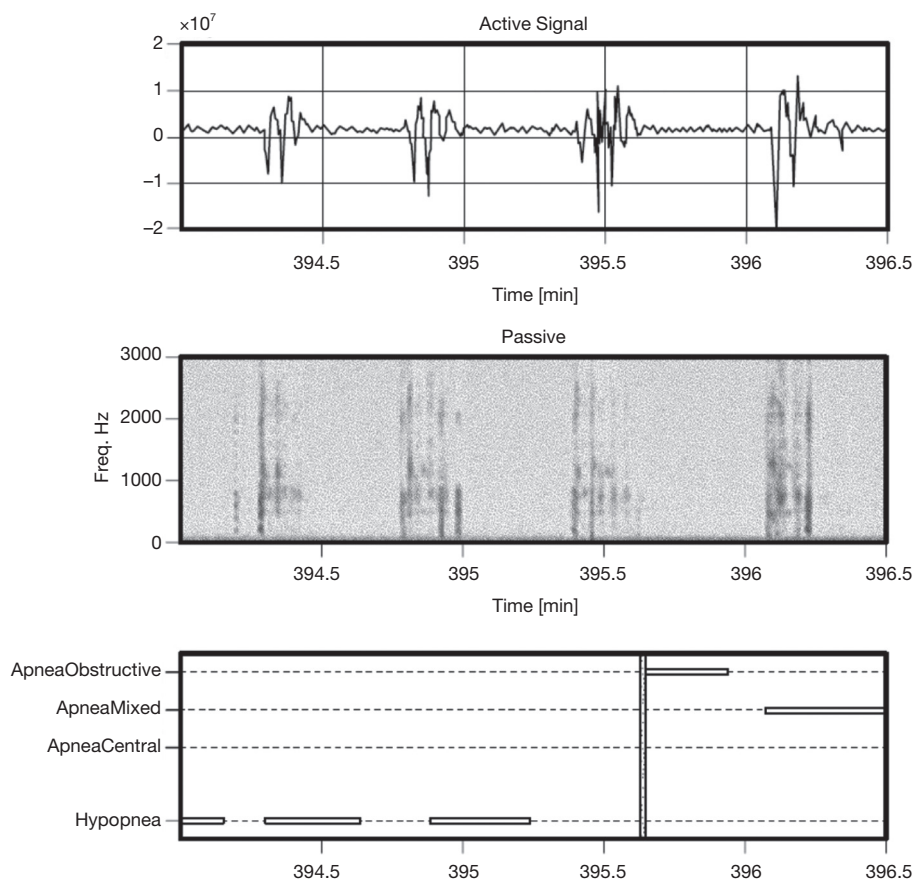
For the screener, a method that provides a binary classification (“positive” or “negative”), with an adjustable

working point based on probability and given threshold, is used. This screener is implemented as a logistic model, which is used to estimate the probability of the night being positive (AHI  $\geq 15$  events per hour of actual sleep time). Separately, an AHI number is estimated, which is derived using a linear regression model.

After the recording session has completed, the data processing is completed on the phone, and a result generated. The result includes:

- (I) An estimated AHI number—similar to significant breathing interruptions (of >10 seconds) average per hour, and;
- (II) A binary classification indicating if the number of significant breathing interruptions was  $\geq 15$  events per hour of sleep (detected risk of OSA/“positive”), or <15 per hour of sleep time (low risk of OSA/“negative”).

The app can be tailored to generate a portable document



**Figure 6** Example of hypopneas, for a patient with AHI of 42.2 events per hour. AHI, apnea hypopnea index.

format (PDF), secure email, link to Electronic Health Record (EHR) or other report format that could be shared with a physician/clinician for review, and follow-up if required.

### Protocol

A total of 248 adults (age  $\geq 21$  years) participated in three studies which took place across the following periods of time: Jun–Aug 2017, Feb–Sep 2018, and Apr–Sep 2019 at the Advanced Sleep Research GmbH (ASR), a contract research organization, in Berlin, Germany. These studies were performed in accordance with the protocol and all applicable laws, rules, regulations, guidelines and standards including the provisions of the World Medical Associations Declaration of Helsinki, German MPG and all relevant confidentiality, privacy and security of patient/test subject information laws [Bundesdatenschutzgesetz (BDSG)].

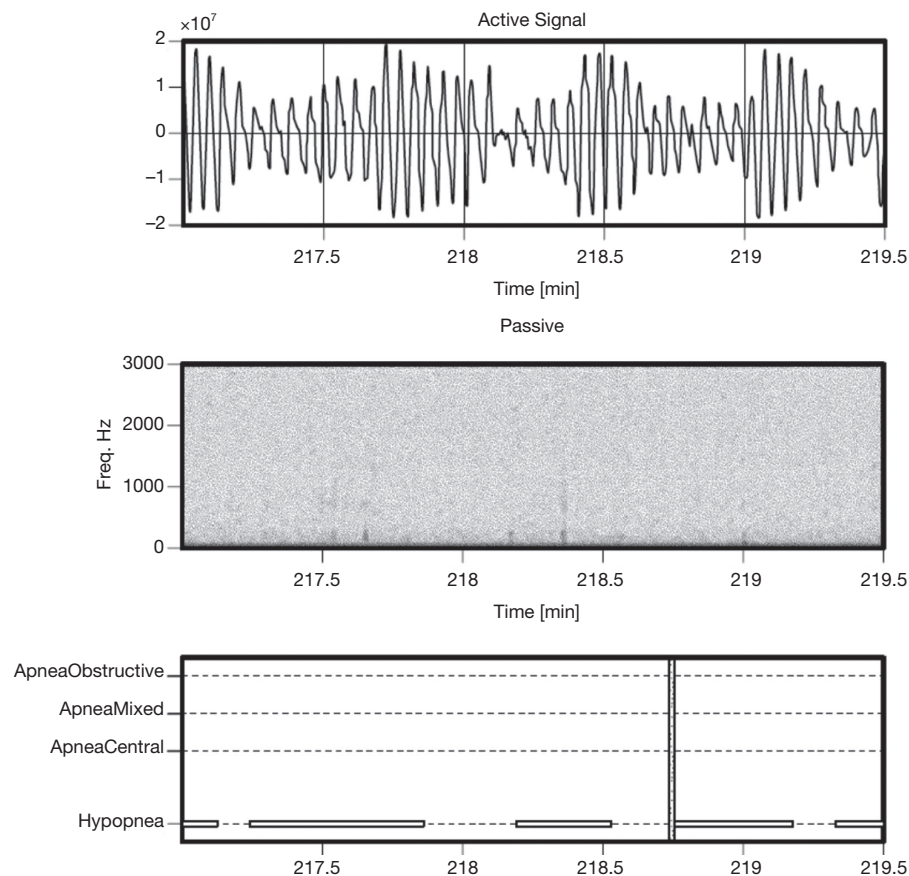
Informed consent to participate was obtained, and subjects were monitored for one night with a full PSG

system and either an Android or an Apple smartphone (Samsung (Seoul, South Korea) S7, Apple (Cupertino, CA, USA) iPhone 7, iPhone 8 Plus or iPhone XS) equipped with a Firefly logging app in a sleep lab setting.

A total of 128 recordings were used for training (S7—100, iPhone7—7, iPhone 8 Plus—17 and iPhone XS—4) while the remaining 120 recordings (S7—74, iPhone7—15, iPhone 8 Plus—19, iPhone XS—12) were kept back for independent testing. In terms of process within ResMed, an algorithm team developed the processing system on the training data, a software team implemented the SDK, and separate quality assurance (QA) and system test teams performed independent assessment of the performance of the released SDK.

The clinical characteristics and demographic information for the training and the test sets are presented in *Table 1*. A breakdown of SDB event types per severity class across the PSG pooled dataset is provided in *Table 2*. As can be seen, hypopneas dominate in terms of SDB events, and thus





**Figure 7** Example of hypopneas, for a patient with AHI of 74.8 events per hour. AHI, apnea hypopnea index.

the authors are confident that the algorithm has adequate training data to recognize this class.

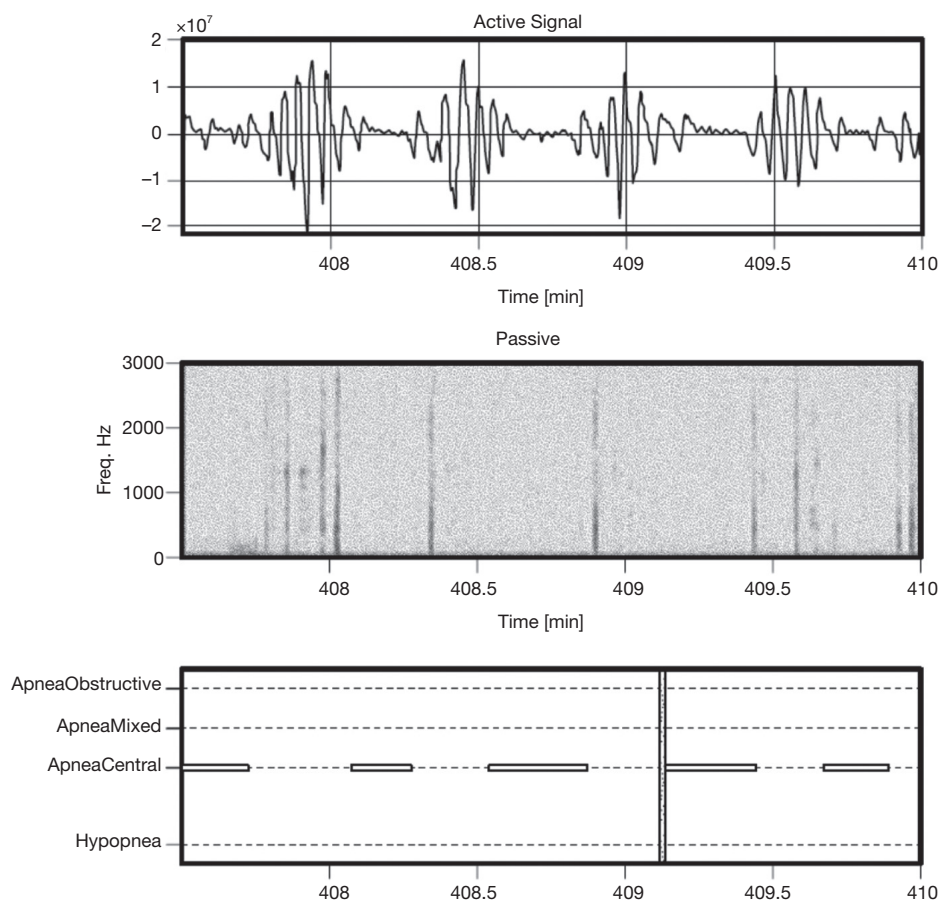
As less than 10% of the subjects have clinically significant levels of central apneas as scored by the PSG lab, the data are not sufficiently powered to allow reporting performance of the algorithm on this subset. However, it is noted that by design the Firefly algorithm is able to detect central apneas and different types of periodic breathing with reasonable accuracy based on the active sonar signal, even in the absence of passive snoring/recovery breaths which typically accompany obstructive events and not central ones. In other words, the sonar stream captures breathing motion, and hence variation in respiratory effort associated with central apneas (e.g., see *Figure 8*: upper panel).

PSG recordings for the first two periods were scored by the same expert scorer using the AASM 2012 guidelines (40). Recordings for the final period were scored by a different scorer using the updated AASM 2018 guidelines (41); the changes in guidelines are not material for this analysis,

as a criteria of 3% desaturation for hypopnea was used in all three periods (this only became mandatory in the 2018 guidelines). The Firefly smartphone recordings were aligned with the PSG recordings by using the timestamps of each device, which were synchronized to an internet time service or cellular/mobile network when available. The QA team also carried out movement analysis on the PSG accelerometer channels, and cross correlated with Firefly movement analysis to confirm alignment.

The test dataset was randomly selected by QA to have an equal proportion of each class. One consequence of this step was a resulting class imbalance (a skew to normal and mild OSA, with fewer moderate and severe cases) in the training dataset as noted in *Table 1*; accordingly, the training process required oversampling to correct for this imbalance.

The logistic model was trained using regularization with five-fold cross validation. The linear regression model used robust fitting of residuals (i.e., to better tolerate outliers that could confound other approaches such as “least squares”).



**Figure 8** Example a subject with central apnea events, AHI of 37.0 events per hour. AHI, apnea hypopnea index.

Due to the relatively sparse data available for training, this approach was used to avoid over-fitting.

As a further processing step, if the estimated AHI value was found to not match the binary classification class, a rescaling of the estimated AHI was carried out; e.g., an estimated AHI via linear regression is say 14.7 events per hour, but the corresponding binary output is “positive” (risk of OSA), the status will remain “positive”, and the estimated AHI rescaled as 15 events per hour.

## Results

### AHI estimation

The results of the linear regression model for the training and withheld test sets of Firefly estimated AHI versus lab human expert scored PSG are illustrated in *Figure 12* as a scatter plot.

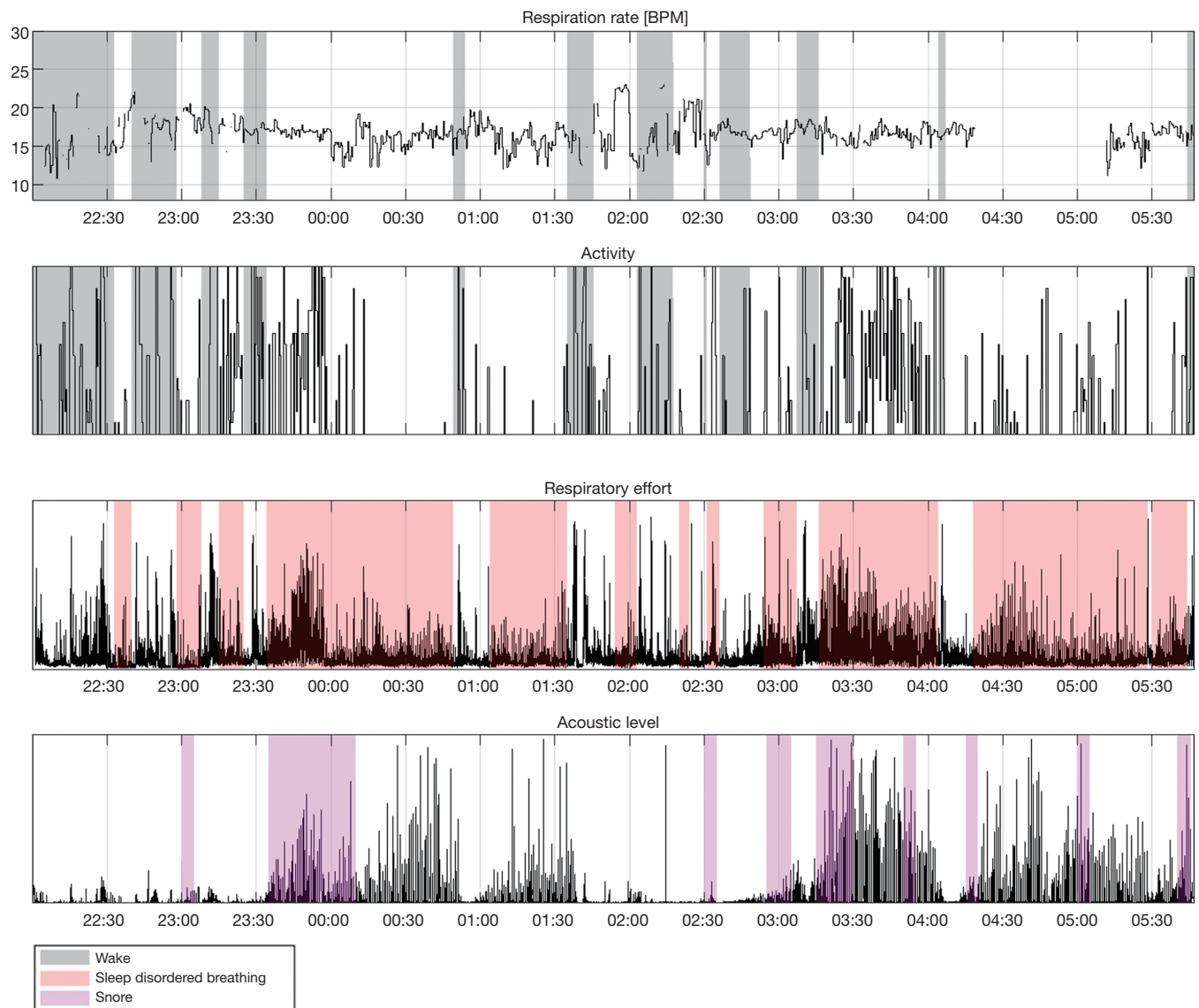
Modified Bland-Altman (42) plots are provided for these

data in *Figure 13*.

### Screeener classification for $AHI \geq 15$ (“negative” or “positive”)

For datasets that are significantly skewed (such as for the training data severity category balance indicated in *Table 1*), precision recall (PR) curves can provide useful insight into model performance (43). Therefore, the performance of the classifier based on the logistic regression model is presented for the training and the testing sets in *Figure 14*, which depicts both the receiver operating characteristic (ROC) and PR curves.

The confusion matrix corresponding to the selected working operating point of the Firefly screener (“0” low risk, “1” high risk of OSA) is shown in *Table 3*. Key performance metrics [sensitivity, specificity, positive predictive value, negative predictive value, accuracy, ROC area under the curve (AUC), and PR AUC] are provided



**Figure 9** A full night view of a Firefly app recording, showing (from top to bottom): estimated respiration rate in breaths per minute, relative activity level, respiratory effort (derived from respiratory envelope), and relative acoustic level (filtered). Wake, SDB, and snore events are overlaid.

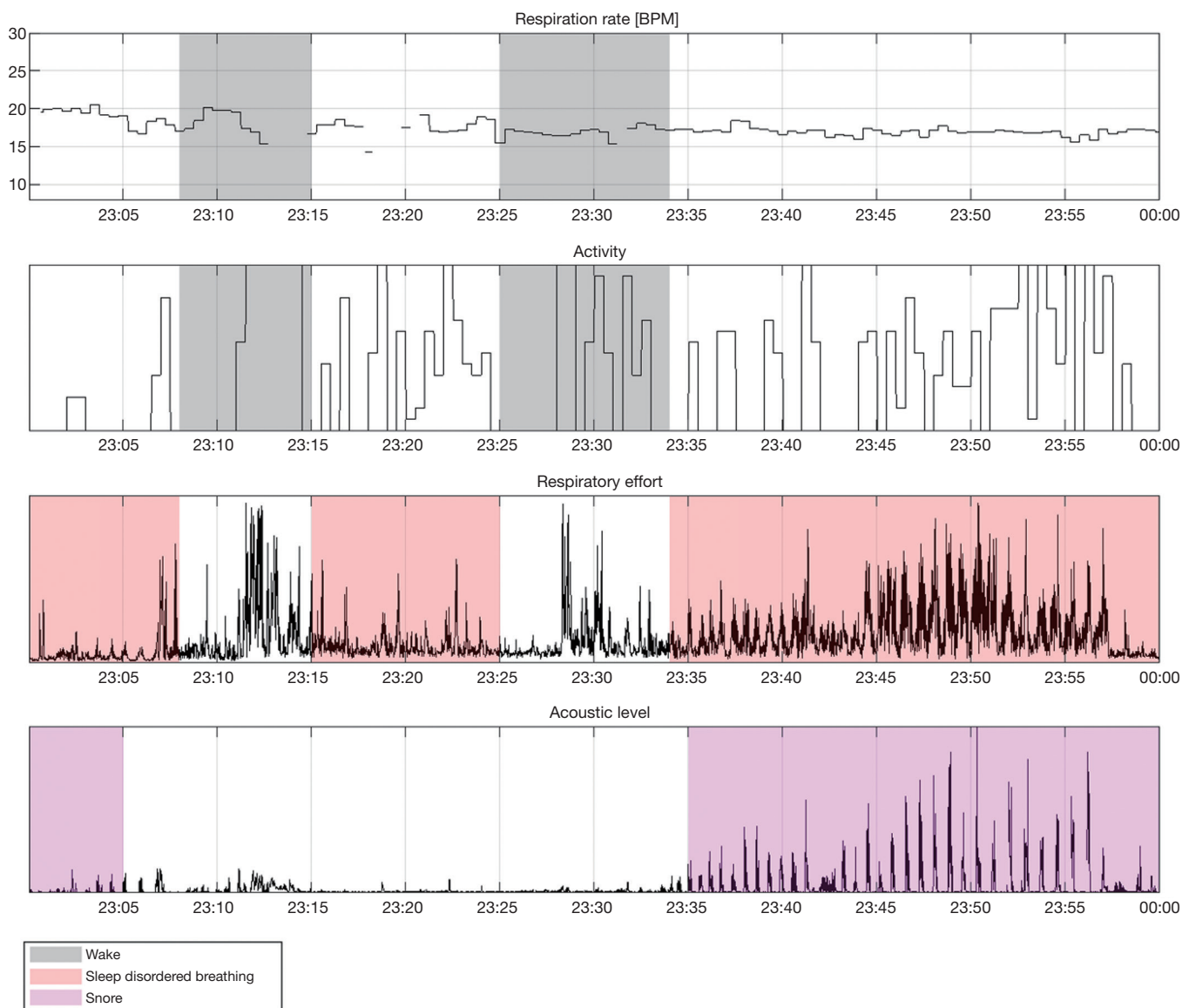
in *Table 4*. A confusion matrix corresponding to the same operating point of Firefly for each OSA class (normal, mild, moderate, severe) is provided in *Table 5*.

#### *Performance comparison with other apps*

A comparison of Firefly to ResApp (44), and to Resonea Drowzle (29) (a phone app OSA screener that recently obtained FDA approval), is provided in *Table 6*. For ROC AUC, the performance of the Firefly system on the

withheld test set exceeds both other systems.

The default operating point (OP) for Firefly was used to compare with ResApp. Drowzle reports two operating points in (29), where the first OP corresponds to a high sensitivity setting (the default configuration of Drowzle), and the second OP has lower sensitivity and higher specificity. It should be noted that the Firefly production OP is set for mid-sensitivity; thus, in order to provide a fair comparison, high and low sensitivity operating points were also processed for Firefly, see *Table 6*. Firefly shows



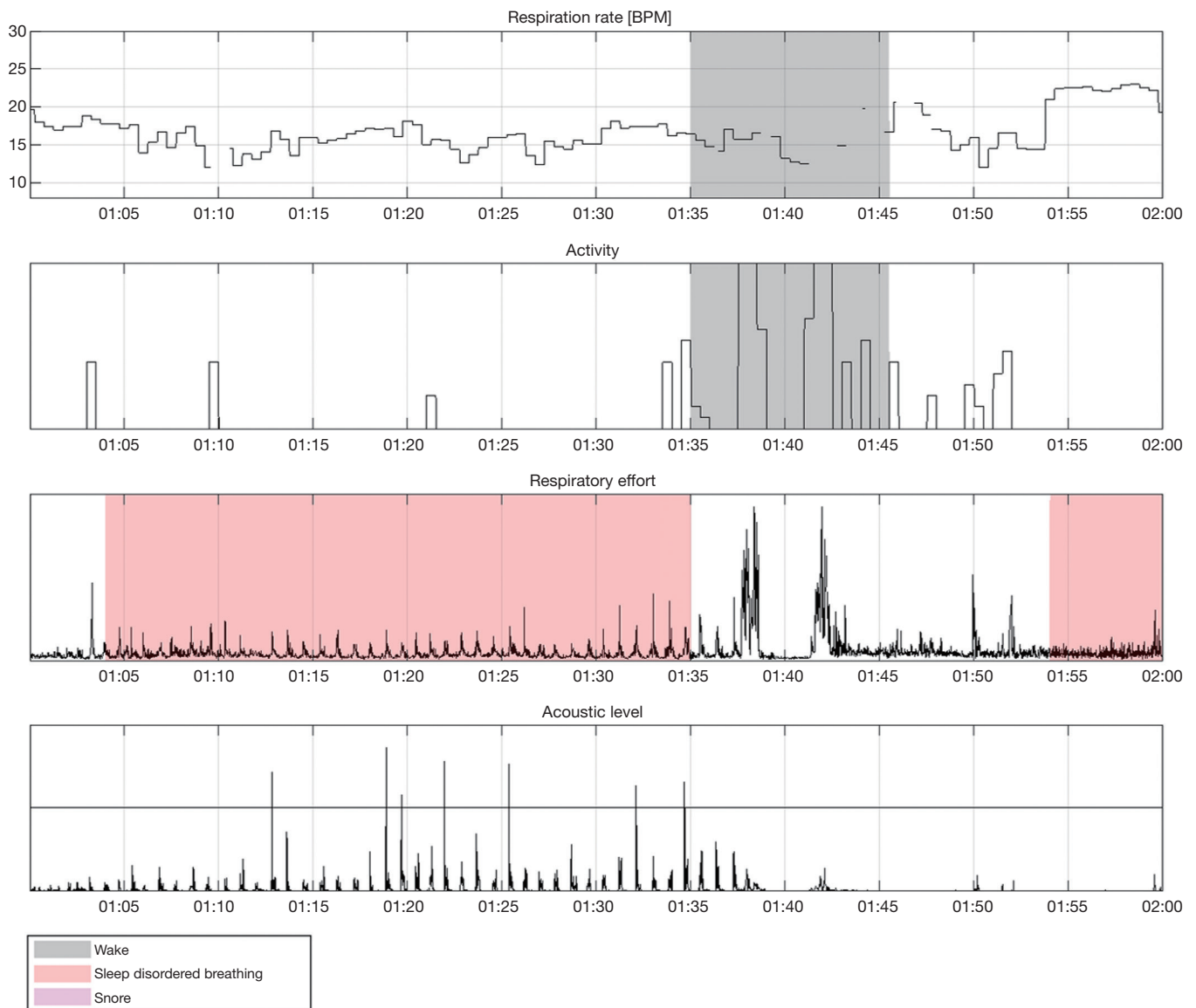
**Figure 10** “Hour 2” segment of the recording depicted in *Figure 9*. In this segment, apneas are detected with associated snoring. The patient woke up twice in this segment.

similar sensitivity and higher specificity than Drowzle for both operating points. Firefly shows similar sensitivity and specificity to ResApp (with a slight bias to higher sensitivity, and slightly reduced specificity).

## Discussion

The dissonance between the high prevalence (and significantly under-diagnosed) condition of OSA, and the existing diagnostic lab PSG pathway (with associated high

costs and long waiting lists), is jarring (1,45). In a world forever changed by COVID-19, it is becoming clear that certain aspects of healthcare will need to evolve to better attend to our post-pandemic needs. Arising from this, the authors believe that a significant opportunity exists for newly evolving contactless, smartphone based approaches to allow large scale screening of OSA. The Firefly technology has the significant benefit of having a low barrier to access, i.e., just requiring the downloading and installing an Android or Apple App on a supported smartphone. An



**Figure 11** “Hour 4” segment of the recording depicted in *Figure 9*. In this segment, apneas are detected without associated snoring.

added benefit is that this can potentially enable scarce and valuable PSG resources to be better focused on complex and comorbid cases.

The Firefly technology platform benefits from a hybrid processing approach, such that two related data streams (active sonar, and passive acoustic) are processed to enhance the robustness of the system. The passive arm calculates biometric features in the audible acoustic frequency range relating to movement, breathing and snoring, whereas the sonar stream measures respiratory effort and movement based on changes in a reflected waveform.

The performance of Firefly on the withheld test set as an OSA screener (with a sensitivity of 88.3% and specificity of 80.0% for a clinical threshold of AHI  $\geq 15$  events/hour—detailed in *Table 4*) has been shown to be comparable to published results of other audio based screeners ResApp (43), and Resonea Drowzle (29).

The system also provides an estimate for the AHI score, which exhibits good performance as depicted in *Table 5* and *Figure 12*, with a correlation of 0.81 on the testing set. Specifically, 61.7% (74/120) of the AHI results from the Firefly technology match their correspondent PSG result in

**Table 1** Demographics and clinical description of the Firefly training and test datasets

Variable	Training set	Test set
Subjects (% male)	128 (52%)	120 (68%)
Age median [mean, SD] (years)	52 [49, 15]	54 [52, 14]
AHI median [mean, SD] (events/hour)	7 [13, 18]	14 [20, 20]
PLM median [mean, SD] (events/hour)	1 [7, 15]	2 [9, 17]
BMI median [mean, SD] (kg/m <sup>2</sup> )	26.3 [27.1, 5.2]	27.6 [28.2, 5.5]
Subjects by severity category		
Normal AHI <5	59 (46%)	30 (25%)
Mild 5 ≤ AHI <15	30 (23%)	30 (25%)
Moderate 15 ≤ AHI <30	20 (16%)	30 (25%)
Severe AHI ≥30	19 (15%)	30 (25%)

AHI, apnea hypopnea index; PLM, periodic limb movement; BMI, body mass index; SD, standard deviation.

**Table 2** PSG pooled data for the 248 recordings used in training and validation: breakdown of SDB event types per severity class

Severity class	# Hypopneas	# Central apneas	# Mixed apneas	# Obstructive apneas
Normal	692	66	6	32
Mild	2,833	174	17	257
Moderate	5,717	408	69	826
Severe	9,034	883	568	3,100

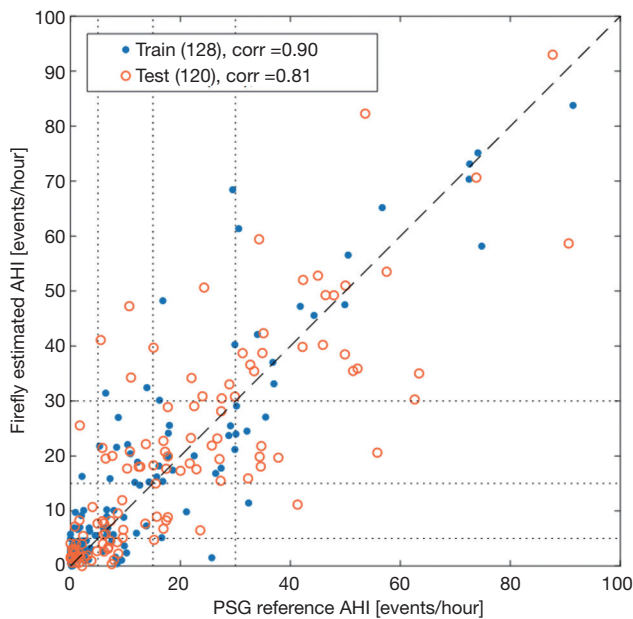
PSG, polysomnography; SDB, sleep-disordered breathing.

the class of OSA severity (none, mild, moderate or severe). Encouragingly, where differences were detected, the errors are limited to the adjacent class of OSA severity in 95% of the cases, with only 6 recordings out of 120 showing difference of OSA severity class prediction larger than one level. The system has good specificity with only two normal subjects being classified as moderate.

Some limitations of this work are as follows. As a single site in Europe was used for lab PSG, the data of the 248 adults will likely cover only a subset of all of the possible manifestations of SDB. These limitations are counterbalanced by using multiple human expert scorers, and also by algorithm design, whereby the robustness of the logistic classification model (which employs a small number of features) considerably reduces the risk of over-fitting. Some degradation in performance is seen between the training and test sets, although this only equated to a drop in accuracy of 3.3%. Not every supported phone handset was tested in the PSG lab; this was counterbalanced by

using several examples of each handset model, and rotating these. Additionally, engineering lab testing of all supported phones is carried out to assess speaker and microphone performance variation, testing with breathing phantoms, and so forth, in order to confirm performance.

Looking more closely at the Bland Altman plots presented in *Figure 13*, it is informative to examine the recordings outside  $\pm 2$  standard deviations (SD) around the median bias. Looking at the false positives (FP) (i.e., AHI <15 on PSG, but Firefly predicted as  $\geq 15$ ), four recordings exhibit severe periodic limb movements during sleep (PLMS). A table looking at PLMS  $\geq 15$  events per hour in the full dataset (N=248) is provided in *Table 7* which shows 9 FP with PLMS  $\geq 15$  (out of 26 FP, equating to 35% of all false positives); 8 of those 9 have already an underlying SDB condition, with a median PSG AHI of 10.7 events per hour. For the remaining false positives, the most likely mechanism is the presence of respiratory events that resemble hypopneas but do not meet the



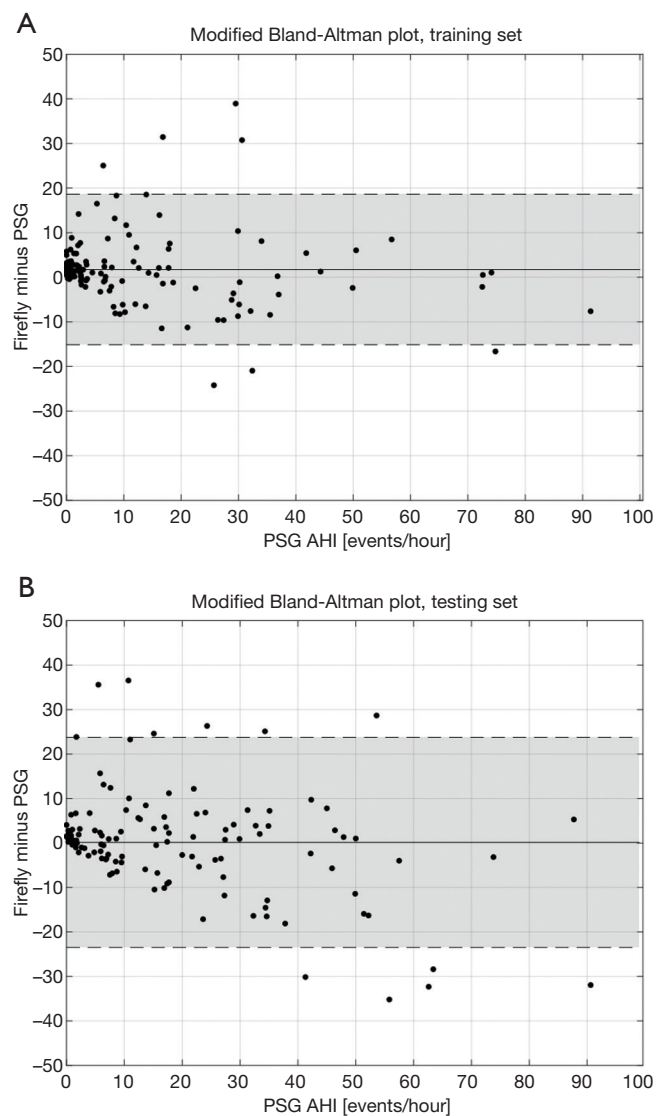
**Figure 12** Scatter plot of the AHI values estimated by Firefly versus the PSG reference, for the training and testing set respectively. The Pearson correlation coefficient is 0.90 on the training set, and 0.81 on the test set. AHI, apnea hypopnea index; PSG, polysomnography.

specific criteria, or acoustic or movement interferers that are similar to the apnea/hypopnea modulation time, and falsely flag OSA.

In terms of false negatives (FN) (i.e., AHI  $\geq 15$  on PSG, but Firefly predicted as  $< 15$ ), potential factors include a presentation of OSA (particularly hypopneas) that does not give rise to the typical envelope modulation morphology; other factors include the subject knocking into the smartphone (or the charging cable) with the duvet (comforter), and in one recorded case managing to cover the phone with the duvet; it is noted that this is more likely in unfamiliar lab/sleeping surroundings, and when instrumented with PSG equipment.

BMI is reasonably well balanced between the training and test sets (as per *Table 1*), and when inspecting the results did not appear to be a factor in the FP or FN cases. Indeed, lab bench testing included a range of BMIs and associated breathing patterns (such as patterns typical in obese subjects) during the design process such as to reduce variability.

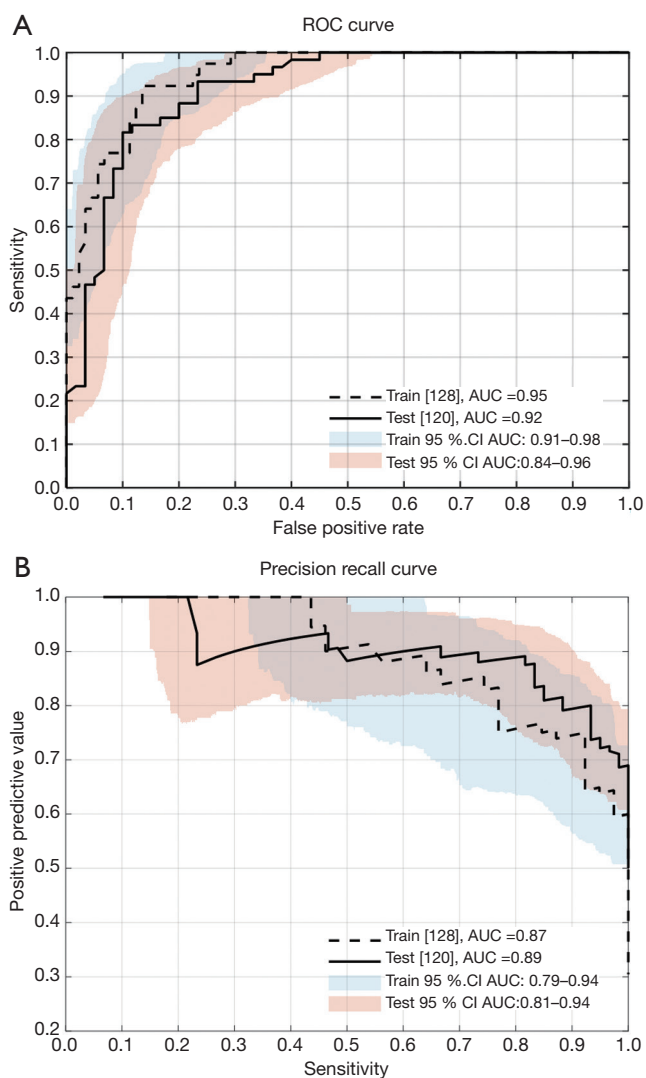
Sleeping body position variability (supine, prone, left, right, different angles to smartphone and so forth) was



**Figure 13** Modified Bland-Altman plots of Firefly estimated AHI versus the reference PSG values for the training set and testing set (A, B respectively). The continuous line corresponds to the median bias, and the shaded areas denote  $\pm 2$  standard deviations (SD) around the median bias. AHI, apnea hypopnea index; PSG, polysomnography.

analyzed and incorporated at the design phase to ensure robustness to body position change throughout the night in an unconstrained setting, and was not separately considered during analysis of the PSG data (which is a constrained setting, as the subjects were instrumented).

It is noted that the constrained nature of the laboratory setting in ASR might not be fully representative of the



**Figure 14** Upper panel (A) presents the receiver operating characteristic (ROC) curves, and panel (B) the precision recall (PR) curves for the logistic classifier, using a clinical threshold of AHI  $\geq 15$ , for the training and testing sets (including 95% confidence intervals). AHI, apnea hypopnea index; AUC, area under the curve; CI, confidence interval.

variability of setup conditions that can be encountered in a home environment. However, some variability in room type was provided, with different sizes and configuration, as well as air conditioning being used on some nights (a potential interferer with other acoustic systems). Furthermore, ResMed has prior experience in deploying contactless sleep measurement in the home environment in unsupervised conditions, and managing the associated variability, such

**Table 3** Classification table for the Firefly technology as a screener for a clinical threshold of AHI  $\geq 15$  versus the PSG classification

Firefly technology	PSG classification	
	Negative	Positive
Negative	Train: 77, test: 48	Train: 4, test: 7
Positive	Train: 12, test: 12	Train: 35, test: 53

AHI, apnea hypopnea index; PSG, polysomnography.

**Table 4** Performance of the Firefly technology as a screener for a clinical threshold of AHI  $\geq 15$

Performance (AHI $\geq 15$ )	Training set	Test set
Sensitivity (%) (CI)	89.7 (75.4–97.3)	88.3 (77.3–95.1)
Specificity (%) (CI)	86.5 (77.9–92.3)	80.0 (67.8–88.7)
PPV (%) (CI)	74.5 (59.5–85.7)	81.5 (70.9–89.6)
NPV (%) (CI)	95.1 (88.5–98.7)	87.3 (76.1–94.4)
Accuracy (%) (CI)	87.5 (81.3–92.3)	84.2 (75.8–89.2)
ROC AUC (CI)	0.95 (0.91–0.98)	0.92 (0.85–0.95)
PR AUC (CI)	0.87 (0.78–0.94)	0.89 (0.80–0.94)

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and AUC of the ROC and PR curves are provided for the training and testing sets respectively. The 95% confidence intervals are shown in parenthesis. AHI, apnea hypopnea index; ROC, receiver operating characteristic; AUC, area under the curve; PR, precision recall.

as with the S+ by ResMed RF and SleepScore Max bedside devices (46).

Work by Stöberl *et al.* suggests that there is high night-to-night variability of OSA, and that the probability of missing moderate OSA was up to 60%; they noted that single-night diagnostic sleep studies are prone to mis-categorize OSA if arbitrary thresholds are used (47). Thus, technology such as Firefly that does not interfere with the user's sleeping habits, and is easy to use for multiple nights offers the ability for longitudinal monitoring of subjects in their home environment in order to more reliably identify the condition.

Other potential benefits of the technology include the simultaneous assessment of sleep quality and fragmentation [such as discussed in (33)] and respiratory rate, snoring, and OSA change on and off PAP therapy. This could provide new insights for physicians into the response to PAP treatment, and increase patient long term adherence.



**Table 5** Classification table for the Firefly technology AHI score in terms of the PSG severity classes (normal AHI <5, mild 5 ≤ AHI <15, moderate 15 ≤ AHI <30 and severe AHI ≥30). Results on the training and testing sets are shown separately

Firefly SDK	PSG classification			
	Normal	Mild	Moderate	Severe
Normal	Train: 48, test: 24	Train:7, test: 9	Train: 1, test: 1	Train: 0, test: 0
Mild	Train: 10, test: 5	Train: 12, test: 10	Train: 2, test: 5	Train: 1, test: 1
Moderate	Train: 1, test: 1	Train: 9, test: 8	Train: 13, test: 17	Train: 4, test: 6
Severe	Train: 0, test: 0	Train: 2, test: 3	Train: 4, test: 7	Train: 14, test: 23

AHI, apnea hypopnea index; PSG, polysomnography; SDK, software development kit.

**Table 6** Comparing performance of Firefly with Drowzle and ResApp

Performance (AHI ≥15)	Firefly training	Firefly testing	Drowzle	ResApp
ROC AUC	0.95	0.92	0.87	0.91
High sensitivity OP (Sens%, Spec%)	(94.9, 77.5)	(93.3, 70.0)	(93.7, 63.0)	N/A
Selected Firefly mid sensitivity OP (Sens%, Spec%)	(89.7, 86.5)	(88.3, 80.0)	N/A	(86.0, 83.0)
Low sensitivity OP (Sens%, Spec%)	(76.9, 92.1)	(81.7, 90.0)	(78.1, 85.2)	N/A

AHI, apnea hypopnea index; ROC, receiver operating characteristic; AUC, area under the curve; OP, operating point.

**Table 7** Periodic limb movements during sleep (PLMS) details across training and test sets

SDB category	n	PLMS ≥15 events per hour	Prevalence PLMS ≥15 events per hour	FP	FP with PLMS ≥15 events per hour*
Normal	89	8	9%	2	1
Mild	60	13	22%	24	8
Moderate	50	12	24%	NA	NA
Severe	49	9	18%	NA	N/A

\*Median AHI for FP with PLMS ≥15 events per hour is 10.7 events per hour. SDB, sleep-disordered breathing; FP, false positive detection.

## Conclusions

The results demonstrate that the Firefly App and SDK technology performs both reliably and accurately in the detection of clinically significant OSA, and in the estimation of AHI, when compared to a PSG gold standard. It offers significant benefits such as the ability to run on Apple and Android smartphones without any added hardware, provides integrated sleep staging, respiration rate analysis, and the ability to monitor for multiple nights in the home environment.

## Acknowledgments

The authors would like to thank the volunteers that took

part in the study, and also Stephen Dodd, Ehsan Chah, Nicola Caffrey & Alberto Zaffaroni (ResMed), and Maria Maas (ASR) for their valuable help.

*Funding:* None.

## Footnote

*Provenance and Peer Review:* This article was commissioned by the editorial office, *Journal of Thoracic Disease* for the series “Sleep Section”. The article was sent for external peer review

*Data Sharing Statement:* Available at <http://dx.doi.org/10.21037/jtd-20-804>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jtd-20-804>). The series “Sleep Section” was commissioned by the editorial office without any funding or sponsorship. RS and SMcM have patents WO/2020/104465-Methods and Apparatus for Detection of Disordered Breathing, WO/2019/122412-Apparatus, System, And Method for Health And Medical Sensing, WO/2019/122414-Apparatus, System, and Method for Physiological Sensing In Vehicles, US 201662396616 Apparatus, System, And Method for Detecting Physiological Movement from Audio and Multimodal Signals, and WO2019122413-Apparatus, System, and Method for Motion Sensing pending. RT and GL have patents WO/2020/104465 and US 201662396616 pending. MW has a patent WO/2019/122412 and WO2019122413 pending. NF and NOM have patent US 201662396616 pending. IF reports grants from Löwenstein, grants from Philips, personal fees from ResMed, outside the submitted work. TP reports grants from ResMed, during the conduct of the study; grants from ResMed, grants and personal fees from Philips, grants and personal fees from Löwenstein Medical, personal fees from Jazz Pharma, personal fees from Heel Pharma, grants from Itamar Medical, other from Bayer Healthcare, outside the submitted work; and Shareholder with Advanced Sleep Research GmbH, Somnico GmbH, and The Siestagroup GmbH. The authors have no other conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The three pilot studies (2017-2019) which included the collection of data from an acoustic smartphone “App” against PSG were performed by Advanced Sleep Research GmbH (ASR), a contract research organization, having its principal place of business at Luisenstraße 54/55, 10117 Berlin, Germany with ResMed. We performed these studies in accordance with the protocol and all applicable laws, rules, regulations, guidelines and standards including ICH GCP guidelines, the provisions of the World Medical Associations Declaration of Helsinki (as revised in 2013), German MPG and all relevant confidentiality, privacy and security of patient / test subject information laws (Bundesdatenschutzgesetz). The exception to compliance with GCP was the lack of submission to and approval by an independent ethics committee. The rationale

for this decision was based on (I) the researchers being comfortable because informed consent was obtained from participants, (II) the protocol was checked prior to the study commencement, and (III) ethics is not required for mobile software/app technology as it was not considered a medical device during the time of the clinical trial, but a consumer device and software and had no intended change of care or intervention to a person/patient.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med* 2019;7:687-98.
2. Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: A population health perspective. *Am J Respir Crit Care Med* 2002;165:1217-39.
3. McNicholas WT, Luo Y, Zhong N. Sleep apnoea: A major and under-recognised public health concern. *J Thorac Dis* 2015;7:1269-72.
4. Epstein LJ, Kristo D, Strollo PJ, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 2009;5:263-76.
5. Osman AM, Carter SG, Carberry JC, et al. Obstructive sleep apnea: current perspectives. *Nat Sci Sleep* 2018;10:21-34.
6. Donovan LM, Patel SR. Making the most of simplified sleep apnea testing. *Ann Intern Med* 2017;166:366-7.
7. Alshaer H, Fernie GR, Tseng WH, et al. Comparison of in-laboratory and home diagnosis of sleep apnea using a cordless portable acoustic device. *Sleep Med* 2016;22:91-6.
8. Norman MB, Pithers SM, Teng AY, et al. Validation of the sonomat against PSG and quantitative measurement of partial upper airway obstruction in children with sleep-disordered breathing. *Sleep* 2017;40(3).

9. Reichert JA, Bloch DA, Cundiff E, et al. Comparison of the NovaSom QSG, a new sleep apnea home-diagnostic system, and polysomnography. *Sleep Med* 2003;4:213-8.
10. Collop NA, Tracy SL, Kapur V, et al. Obstructive sleep apnea devices for Out-Of-Center (OOC) testing: Technology evaluation. *J Clin Sleep Med* 2011;7:531-48.
11. Michaelson PG, Allan P, Chaney J, et al. Validations of a portable home sleep study with twelve-lead polysomnography: comparisons and insights into a variable gold standard. *Ann Otol Rhinol Laryngol* 2006;115:802-9.
12. Penzel T, Schöbel C, Fietze I. New technology to assess sleep apnea: Wearables, smartphones, and accessories. *F1000Res* 2018;7:413.
13. De Chazal P, Fox N, O'Hare E, et al. Sleep/wake measurement using a non-contact biomotion sensor. *J Sleep Res* 2011;20:356-66.
14. Hong H, Zhang L, Gu C, et al. Noncontact Sleep Stage Estimation Using a CW Doppler Radar. *IEEE J Emerg Sel Top Circuits Syst* 2018;8:260-70.
15. Park KS, Choi SH. Smart technologies toward sleep monitoring at home. *Biomed Eng Lett* 2019;9:73-85.
16. O'Hare E, Flanagan D, Penzel T, et al. A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. *Sleep Breath* 2015;19:91-8.
17. Ballal T, Shouldice RB, Heneghan C, et al. Breathing rate estimation from a non-contact biosensor using an adaptive IIR notch filter. *RWW 2012-Proc. 2012 IEEE Top. Conf. Biomed. Wirel. Technol. Networks, Sens. Syst. BioWireless* 2012;2012:5-8.
18. Ballal T, Heneghan C, Zaffaroni A, et al. A pilot study of the nocturnal respiration rates in COPD patients in the home environment using a non-contact biomotion sensor. *Physiol Meas* 2014;35:2513-27.
19. Shouldice RB, Heneghan C, Petres G, et al. Real time breathing rate estimation from a non contact biosensor. *Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf* 2010;2010:630-3.
20. Kang S, Lee Y, Lim YH, et al. Validation of noncontact cardiorespiratory monitoring using impulse-radio ultra-wideband radar against nocturnal polysomnography. *Sleep Breath* 2020;24:841-8.
21. Zaffaroni A, De Chazal P, Heneghan C, et al. SleepMinder: An innovative contact-free device for the estimation of the apnoea-hypopnoea index. *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009, IEEE Computer Society, 2009:7091-4.*
22. Zaffaroni A, Kent B, O'Hare E, et al. Assessment of sleep-disordered breathing using a non-contact bio-motion sensor. *J Sleep Res* 2013;22:231-6.
23. Crinion SJ, Tiron R, Lyon G, et al. Ambulatory detection of sleep apnea using a non-contact biomotion sensor. *J Sleep Res* 2020;29:e12889.
24. Xiong J, Jiang J, Hong H, et al. Sleep apnea detection with Doppler radar based on residual comparison method. 2017 International Applied Computational Electromagnetics Society Symposium (ACES). China, ACES-China 2017, 2017.
25. Stryjak J, Sivakumaran M. GSMA Intelligence — Research — The Mobile Economy 2019 n.d. Available online: <https://www.gsmainelligence.com/research/2019/02/the-mobile-economy-2019/731/> (accessed January 9, 2020).
26. Holst A. Number of smartphone users worldwide 2014-2020 | Statista. Statista 2019:1. Available online: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> (accessed January 9, 2020).
27. Poushter J. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies | Pew Research Center. *Pewglobal* 2016:1-9. Available online: <https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies> (accessed January 2, 2020).
28. GSM Association. The Mobile Economy 2019. GSMA 2019:56.
29. Narayan S, Shivdare P, Niranjana T, et al. Noncontact identification of sleep-disturbed breathing from smartphone-recorded sounds validated by polysomnography. *Sleep Breath* 2019;23:269-79.
30. Akhter S, Abeyratne UR, Swarnkar V, et al. Snore sound analysis can detect the presence of obstructive sleep apnea specific to NREM or REM sleep. *J Clin Sleep Med* 2018;14:991-1003.
31. Nandakumar R, Gollakota S, Watson N. Contactless sleep apnea detection on smartphones. *MobiSys 2015-Proc. 13th Annu. Int. Conf. Mob. Syst. Appl. Serv., Association for Computing Machinery, Inc, 2015:45-57.*
32. Lyon G, Tiron R, Zaffaroni A, et al. Detection of Sleep Apnea Using Sonar Smartphone Technology. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Institute of Electrical and Electronics Engineers (IEEE), 2019:7193-6.
33. Zaffaroni A, Coffey S, Dodd S, et al. Sleep Staging Monitoring Based on Sonar Smartphone Technology. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Institute of Electrical and Electronics Engineers (IEEE), 2019:2230-3.
34. McMahan S, O'Rourke D, Zaffaroni A, et al. Apparatus, System, And Method For Detecting Physiological

- Movement From Audio And Multimodal Signals. US 201662396616, 2016.
35. Shouldice R, McMahon S, Wren M. Apparatus, System, And Method For Health And Medical Sensing. US 201762610033, 2017.
  36. Shouldice R, McMahon S. Apparatus, System, and Method For Physiological Sensing In Vehicles. US 201762609998, 2017.
  37. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. GPU Comput Gems Emerald Ed 2016. doi: 1603.04467:277-91.
  38. Heneghan C, Chua CP, Garvey JF, et al. A portable automated assessment tool for sleep apnea using a combined holter-oximeter. *Sleep* 2008;31:1432-9.
  39. Hayano J, Watanabe E, Saito Y, et al. Diagnosis of sleep apnea by the analysis of heart rate variation: A mini review. *Conf Proc IEEE Eng Med Biol Soc* 2011;2011:7731-4.
  40. Berry RB, Budhiraja R, Gottlieb DJ, 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* 2012;8:597-619.
  41. Berry RAC. AASM Scoring Manual version 2.5–American Academy of Sleep Medicine. 2018.
  42. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
  43. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. *ACM Int Conf Proceeding Ser* 2006;148:233-40.
  44. Partridge N, May J, Peltonen V, et al. Large sample feasibility study showing smartphone-based screening of sleep apnoea is accurate compared with polysomnography. *J Sleep Res* 2018;27:e143\_12766.
  45. Arnardottir ES, Bjornsdottir E, Olafsdottir KA, et al. Obstructive sleep apnoea in the general population: Highly prevalent but minimal symptoms. *Eur Respir J* 2016;47:194-202.
  46. Schade MM, Bauer CE, Murray BR, et al. Sleep validity of a non-contact bedside movement and respiration-sensing device. *J Clin Sleep Med* 2019;15:1051-61.
  47. Stöberl AS, Schwarz EI, Haile SR, et al. Night-to-night variability of obstructive sleep apnea. *J Sleep Res* 2017;26:782-8.

**Cite this article as:** Tiron R, Lyon G, Kilroy H, Osman A, Kelly N, O'Mahony N, Lopes C, Coffey S, McMahon S, Wren M, Conway K, Fox N, Costello J, Shouldice R, Lederer K, Fietze I, Penzel T. Screening for obstructive sleep apnea with novel hybrid acoustic smartphone app technology. *J Thorac Dis* 2020;12(8):4476-4495. doi: 10.21037/jtd-20-804