

SleepAp: An Automated Obstructive Sleep Apnoea Screening Application for Smartphones

Joachim Behar¹, Aoife Roebuck¹, Mohammed Shahid¹, Jonathan Daly¹, Andre Hallack¹, Niclas Palmius¹, John Stradling², Gari D. Clifford¹, *Senior Member, IEEE*

Abstract—Obstructive Sleep Apnoea (OSA) is a sleep disorder with long term consequences. Long term effects include sleep-related issues and cardiovascular diseases. OSA is often diagnosed with an overnight sleep test called a polysomnogram. Monitoring can be costly with long wait times for diagnosis. In this paper a novel OSA screening framework and prototype phone application are introduced. A database of 856 patients that underwent at-home **polygraphy** was collected. Features were derived from audio, actigraphy, photoplethysmography (PPG) and demographics, and used as the inputs of a support vector machine (SVM) classifier. The SVM was trained on 735 patients and tested on 121 patients. Classification on the test set had an accuracy of up to 92.2% **when classifying subjects as having moderate or severe OSA vs. being healthy or a snorer based on the clinicians' diagnoses**. The signal processing and machine learning algorithms were ported to Java and integrated into the phone application - SleepAp. SleepAp records the body position, audio, actigraphy and PPG signals and implements the clinically validated STOP-BANG questionnaire. It derives features from the signals and classifies the user as having OSA or not using the SVM trained on the clinical database. The resulting software could provide a new, easy-to-use, low-cost and widely available modality for OSA screening.

Index Terms—obstructive sleep apnoea, sleep disorders, mHealth, actigraphy, audio, PPG

I. INTRODUCTION

Obstructive Sleep Apnoea (OSA) is a disorder characterised by periods of breathing cessation (apnoea) and periods of reduced breathing effort (hypopnoea) during sleep due to the complete or partial collapse of the upper airway. This leads to deoxygenation (as there is no air going into the lungs the arterial oxygen levels drop and the carbon dioxide levels rise) and consequent arousal caused by a surge of sympathetic nervous system activity. Long term consequences include sleep-related issues and cardiovascular diseases. OSA is often diagnosed with an overnight sleep study or polysomnogram (PSG) which is considered to be the gold standard for OSA diagnosis. **Subjects fall into a number of categories based on the average number of apnoea and hypopnoea events (AHI) or the average number of oxygen desaturation events (ODI) they have per hour of sleep: healthy/snorer: $AHI/ODI < 5$; mild OSA: $5 \leq AHI/ODI \leq 15$; moderate OSA: $15 < AHI/ODI \leq 30$; severe OSA: $AHI/ODI > 30$.** Monitoring can be costly (between \$1000 and \$1217 [1]) and wait time for diagnosis

can be long. In addition, patients with undiagnosed OSA are high users of healthcare services with expenditure estimated to be almost double that of age and gender matched controls [2]. **OSA is a common disorder affecting anywhere from 2% to 7% of the adult population globally [3]–[9] but with a high variance across ethnic groups (e.g. an incidence of 3% in Australia [3] and up to 5% in India [9] for males).** It is estimated that up to 90% of subjects with the disorder are undiagnosed and untreated [10].

There are many home sleep recording systems that aim to reduce the financial burden of in-hospital PSG. Examples include the type II Sleepscan Netlink Traveller (Bio-Logic Systems, Mundelein, Illinois, USA); the type II Vitaport-4 PSG (TEMEC Instruments, Kerkrade, Netherlands); the type IV Grey Flash (Stowood Scientific Instruments, Oxford, UK); **Watch-PAT (Itamar Medical Ltd, Tel Aviv, Israel); and Nox T3 (Yorba Linda, California, US).** However, patients who may be unfamiliar with medical instrumentation have to set up the equipment by themselves and the data needs to be reviewed by a sleep professional which is costly and time consuming. This could be avoided with an automated diagnosis algorithm.

The recent increase in adoption of smartphones along with the inclusion of high-quality internal sensors has led to the proliferation of sleep screening smartphone applications (apps). The computational platform and sensors provided by such devices have been used for sleep screening by many recently-released apps. Behar *et al.* [11] reviewed more than 40 sleep apps available on the Android market and Apple App Store and concluded that none of the apps were based on clear scientific evidence (with the exception of apps implementing simple sleep questionnaires). The research question addressed in this work is whether a simple, cheap and effective framework can be created using the capabilities provided by a modern smartphone to screen for and/or monitor OSA.

In this work a novel OSA screening framework and associated Android phone application, ‘SleepAp’ are introduced. SleepAp uses internal phone sensors and an external pulse oximeter to record audio, activity, body position and oxygen saturation during sleep. The algorithms implemented in the app are motivated by signal processing and machine learning algorithms validated on a clinical database. Section II-A introduces the clinical database used to motivate the app framework and validate the algorithms while Section I discusses the recording of the signals on the phone. Section III presents the results obtained on the clinical database and the new Android SleepAp. Finally Section IV discusses the phone app, future work and provides a conclusion.

¹ Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, United-Kingdom

² Sleep Unit, Oxford Centre for Respiratory Medicine, Churchill Hospital, Oxford, United Kingdom

Contact: joachim.behar@eng.ox.ac.uk

Manuscript received March 20, 2013; revised April 21, 2012.

TABLE I: Clinical database

		Training set	Test set
Non-OSA	Healthy	130 (18%)	25 (21%)
	Snoring	238 (32%)	19 (16%)
	Mild OSA	89 (12%)	17(14%)
OSA	Moderate OSA	101(14%)	22(18%)
	Severe OSA	177(24%)	38(31%)
Total		735(100%)	121(100%)

II. METHODS

A. Clinical Database

In order to train and validate the signal processing and machine learning algorithms to be ported to the smartphone platform, 856 randomly chosen patients from the Respiratory Medicine Unit (Churchill Hospital, Oxford, UK) were recorded in a home environment using the Grey Flash home polygraphy recording device (Stowood Scientific Instruments Ltd, Oxford, UK). Patient records were collected over a three year period (2010 to 2013). Overall, the recordings of 1354 subjects were collected, however they had been referred to the respiratory medicine unit and so many of the diagnoses were not relevant (e.g. lung cancer, hypoventilation) or the recordings failed (technical failure or recording less than four continuous hours). Further details of the database used in this analysis can be found in [12]. Each subject's recording was analysed by an experienced clinician and a diagnosis given based on the ODI, sound recording, sleep disturbance from movement, heart rate, pulse transit time (PTT), and respiratory effort from PTT. It should be noted that there can be considerable sleep disturbance even when the ODI is low, for example in upper airway resistance syndrome, so that a subject with an ODI of 7 could be classified as moderate if there are many snoring related arousals as well. Therefore the classification is based on the ODI but modified according to the medical expert's reading/interpretation. Thus true numerical thresholds are not optimal and in the sleep clinic, where the data were recorded and diagnosed, they are regarded as less useful than a more global interpretation. A more complete description of the diagnosis process can be found in Craig *et al.* [13]. Patients were randomly separated into training and test sets (Table I). Features were derived from the audio, actigraphy and photoplethysmogram (PPG) signals and a support vector machine (SVM) classifier was trained on 735 patients (50.1% non-OSA and 49.9% OSA) and tested on 121 patients (36.4%-63.6% split). This experiment is denoted EXP-1. In a second run, denoted EXP-2, the SVM was trained and tested on the data set excluding the mild OSA patients.

B. Framework motivation

The ultimate goal of the developed phone application is to provide an accurate screening tool for OSA by providing the probability of a user belonging to one of the following two classes: non-OSA (healthy and snorers) and OSA (mild, moderate and severe). Audio, actigraphy, body position and PPG can be recorded on the phone and sleep-related features extracted from those signals. Using the sleep-related features in combination with the answers from a clinically validated

TABLE II: List of features

Symbol	Definition	#features	Mean/Std
MSE	from audio	9	-
MSE	from actigraphy	9	-
ODI	from PPG	1	-
BMI	body mass index	1	33.1±9.4
Weight	in kg	1	99.3±27.6
Height	in meters	1	1.7±.1
Age	in years	1	49.4±13.5
Neck size	in centimeters	1	42±4.8
Gender	Male/Female	1	585/271
Total		25	

MSE: multi scale entropy, ODI: oxygen desaturation index, PPG: photoplethysmography, BMI: body mass index, Std: standard deviation.

questionnaire the patients are classified as belonging to one of the two classes. Fig. 1 gives an example of the signals of interest for a patient who was diagnosed with severe OSA (a) and for a patient who was diagnosed as non-OSA (b).

1) *Features*: Many alternative algorithms for deriving features from sleep time series exist (see [14] for a review). Based on the work of Roebuck *et al.* [15], Behar *et al.* [16] and Higgins *et al.* [17], multiscale entropy (MSE) which corresponds to sample entropy calculated over varying time scales was chosen to quantify the irregularity in the audio and actigraphy recordings with extremes of entropy indicating pathological physiology. In this work, using the entropy computed at multiple scales improved classification between OSA and non-OSA individuals when compared to using only one scale for entropy. As such, multiple MSE-based features were used for audio and actigraphy (See Tab. II). MSE was implemented using Physionet's open-source C code [18]. The ODI (derived from the PPG signal) and available demographics were used as features for the SVM classifier. Table II lists all the features.

2) *Machine learning*: We used an SVM classifier (libSVM library [19], [20]) with a Gaussian (non-linear) kernel with two parameters: C (controls the trade-off between minimizing training errors and the model complexity [21]) and γ (width of the Gaussian kernel). Based on cross-validation performed on the training data set, C and γ were optimised for each combination of features using a grid search. The decision boundary of the resulting SVM classifier corresponds to $y(\mathbf{x}) = 0$, where \mathbf{x} is the feature vector and the classification confidence increases with respect to the magnitude of the decision function y . As a consequence, the pairwise probability (i.e. the probability of an occurrence \mathbf{x} belonging to class 1 knowing that $i \in \{-1, 1\}$) is calculated using a sigmoid function whose output value is in $[0 : 1]$ and with the decision boundary corresponding to the point where the sigmoid function is equal to 0.5. Outputting a probability estimate (i.e. the likelihood of having OSA) instead of performing a classification based on decision value is preferable in the context of the app as it will be used for screening purposes. For the statistics reported in this paper the decision threshold is taken to be 0.5. All features were normalised by subtracting the mean and dividing by the standard deviation computed on the training set data. The probability estimate computation is described in [22].

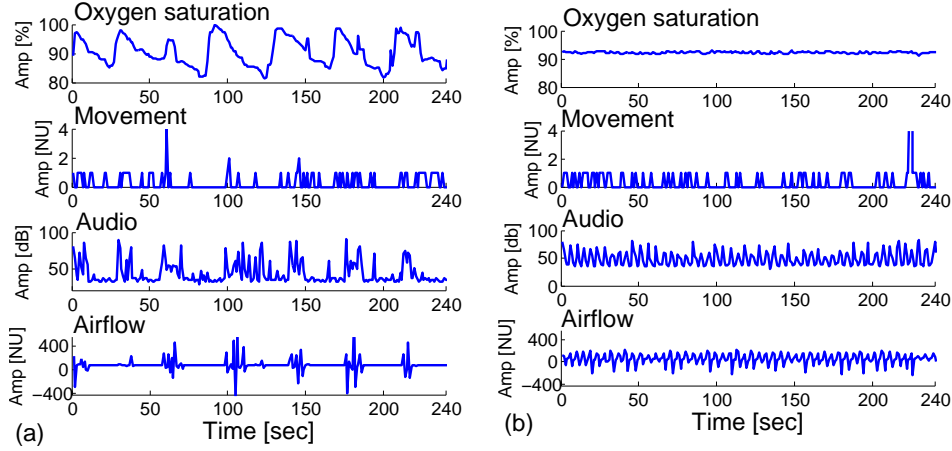


Fig. 1: Illustration of the signals of interest as well as the airflow signal from the clinical database for (a) a patient with severe OSA; and (b) a non-OSA patient. OSA events can be clearly identified on the plots for the patient with severe OSA; multiple drops in oxygen saturation by more than 3-4% of the baseline, no airflow for long intervals and no breathing sounds.

C. Operating system and general set-up

There are a number of suitable mobile phone operating systems (OS) available with the most common including Android, iOS, Windows Phone and BlackBerry. As pointed out by Behar *et al.* [11] the choice of the OS is subject to two criteria: (1) how widespread it is and (2) the homogeneity of the hardware. Android is the best choice considering the first criterion given its large market share but iOS better satisfies the second. For this work we decided to work with Android (versions 2.3.3 to 4.2.2, covering over 95% of Android devices [23]) due to its open source licensing and recent growth in market share, particularly in developing countries [24]. However, there are challenges due to the non-homogeneous hardware which will need to be evaluated further.

D. Recorded signals

1) *Actigraphy*: Actigraphy consists of monitoring human activity. Its measurement during sleep has been shown to be a valid tool in helping to differentiate healthy adult patients from those suffering from some sleep disorders [25]. Accelerometers in mobile phones can be used to record the patient's movement over the night; these sensors come with a variety of different precisions and sensitivities. The framework described here contains an actigraphy analysis module which extracts movement information and transforms it into features which can be used to assess the likelihood of the patient having OSA. The app samples motion at 4 Hz which is in the range of some devices currently used in medical practice for diagnosis of OSA (e.g. 1 Hz and 4 Hz for VISI and Grey Flash respectively, both from Stowood Scientific Instruments Ltd, Oxford, UK).

The accelerometer measures the acceleration applied to the device (\vec{a}). Conceptually, this was done by measuring forces applied to the sensor (\vec{F}_i) with $\sum_{i=1}^N \vec{F}_i = m\vec{a}$, where N is the number of external forces applied to the system, m is the mass of the device. Gravity always influences the measured acceleration; thus, we may rewrite it as: $\sum_{i=1 \neq \text{gravity}}^N \vec{F}_i + m\vec{g} = m\vec{a}$. This phenomenon can be observed by leaving the phone on a flat surface ($\forall i \in N \neq \text{gravity}, \vec{F}_i = 0$) where

the accelerometer gives a value of magnitude $g = 9.8ms^{-2}$. To access the acceleration of the device caused only by the patient's movement the contribution of the force of gravity needs to be removed. This can be achieved by applying a high-pass filter to the accelerometer time-series. A finite impulse response band-pass filter was implemented in Java for filtering the raw data on the phone (with cut-off frequencies at 0.75 Hz and 1.65 Hz). The higher cut-off frequency serves to filter out noise. The discrete model used for the filter is given by

$$w_k[j] = \frac{v_k[j] - v_k[j-1] - v_k[j-2] + v_k[j-3]}{2},$$

where $v_k[j]$ is the input acceleration signal for direction k (from the three cardinal directions x , y and z) and $w_k[j]$ is the filtered acceleration signal. The lower cut-off frequency property of this filter ensures that most of gravity component in the acquired signal is removed. The composite acceleration is then computed as

$$M[j] = \sqrt{w_x^2[j] + w_y^2[j] + w_z^2[j]}.$$

The computed acceleration magnitude signal is grouped every one second and the logarithm of the variance within this second is computed as

$$Y_{ac}(n) = \log(\text{Var}[\overline{M}(n)] + 1),$$

where Y_{ac} is the actigraphy processed signal and $\overline{M}(n) = \{M(t), \forall t \in T : n\Delta T \leq t < (n+1)\Delta T\}$, M is the computed acceleration magnitude signal, t is the signal acquisition time and $\Delta T = 1\text{sec}$. The last processing stage consists of calculating MSE for Y_{ac} .

2) *Body position assessment*: It has been shown that there is a correlation between the severity of apnoeic events and body position during sleep [26]. As a consequence body position could be used as a diagnostic feature during the referral stage. The developed framework measures the proportion of time spent during the night on each of the following positions: supine, sitting, prone and laying sideways (left or right).

Body position can be computed on a mobile phone by taking

Y/X	-180°	-135°	-45°	0°	45°	135°	180°
-90°	2	2	2	2	2	2	2
-45°	1	5	4	4	5	1	1
0°	1	5	4	4	5	1	1
45°	1	5	4	4	5	1	1
90°	3	3	3	3	3	3	3

Fig. 2: Body position with respect to Pitch (X) and Roll (Y) angles. 1=prone, 2=right, 3=left, 4=supine, 5=sitting up.

complimentary measures from two sensors commonly found in smartphones: an accelerometer and a geomagnetic field sensor [27]. SleepAp uses pitch and roll angles to determine in what positions the patient is resting while assuming the phone is worn on an armband oriented parallel to the reference coronal plane. The phone is worn in an armband for convenience. Fig. 2 shows the body position mapping with the pitch and roll angles. The body position was assessed every 0.25 seconds (4 Hz). Each change of the patient's position was recorded and the overall proportion of time spent in each of the different positions overnight was computed. The accelerometer-derived angles can be imprecise in the case of a non-quasi-static system. However for the activity being produced during sleep, these measures are good enough to provide the proportion of time spent in each of the different positions during the night.

3) *Audio*: Recorded audio of the patient's breathing during sleep is an efficient tool for OSA diagnosis. Snoring sounds from healthy and sleep apnoea patients have significant differences and may be used alone to provide accurate classification between these two groups [28]. It is important, particularly when dealing with low-cost equipment, to ensure that the acquired signals retain their relevant diagnostic features. There are multiple parameters in the audio acquisition workflow that have an impact on the audio quality [11]: frequency response of the phone's sound card; the format of the recorded audio file (ie. the encoding and associated compression algorithms); type/quality of microphone, and the location of the microphone in relation to the subject (e.g. on-body or off-body).

Current sleep apnoea audio recording systems acquire the signal with a broad range of sampling rates (5 kHz [28], 20 kHz [29], 44.1 kHz [30]). OSA patients' snoring sounds present a wide range of frequency response spectra depending on the physical phenomenon causing it. In general, obstruction snores have at least 75% of their energy restricted to frequency components under 2 kHz [31]. Since memory is a limited resource when performing recordings with smartphones, audio is acquired in the proposed framework with a sampling rate of 8 kHz. The signal is recorded using uncompressed single-channel (mono) 16-bit pulse-code modulation. The microphone is attached to the face using surgical tape, plasters or any other medical adhesive. The acquired audio signal is processed in a similar way to the accelerometer data (Subsection II-D1).

$$Y_{au}(n) = \log(\text{Var}[\bar{S}(n)] + 1),$$

where Y_{au} is the audio processed signal and $\bar{S}(n) = \{S(t), \forall t \in T : n\Delta T \leq t < (n+1)\Delta T\}$, S is the acquired audio signal, t is the signal acquisition time and $\Delta T = 1\text{sec}$. Finally, MSE is run on Y_{au} .

4) *PPG*: PPG was recorded using the WristOx2 3150 pulse oximeter (Nonin Medical Inc., MN, USA). The pulse oximeter is connected to the app via Bluetooth and records PPG at a sampling frequency of 75 Hz. The WristOx2 is attached to the subject's wrist and one finger (index, middle or ring) is inserted into the finger cuff. The pulse oximeter is powered by two 1.5V AAA batteries and its operating life is given as "24 hours minimum" in the model specification. ODI was derived from the recorded oxygen saturation data with a desaturation event being defined as a decrease of at least 4% from the average oxygen saturation (calculated from the previous 120sec) and lasting for at least 10sec. However, the system is designed to use any pulse oximeter which can communicate with the phone via Bluetooth or USB for example.

E. Feature extraction and classification

Once the features have been derived – from audio, actigraphy, PPG and the questionnaire – the features are classified using a trained SVM. The implementation draws upon the multi-platform open-source software library LIBSVM [32], as Android uses standard Java libraries. Development of the classifier took place in MATLAB (using LIBSVM) to generate a structure containing the SVM type, kernel type, number of classes, total number of support vectors, offset, class labels and the support vectors (see [17] for details). This structure was then saved as a text file to allow the Android application to import it in a standardised format. Since recording PPG requires the user to have a pulse oximeter – which might be relatively expensive or simply unavailable (and thus not realistic for mass screening) – and recording actigraphy requires the user to have an armband, multiple SVM structures were trained on all possible combinations of signals. This means that a prediction can be obtained with any combination of recorded signals (audio, actigraphy, PPG and demographics).

III. RESULTS

A. Results on the clinical database

Tables III and IV present the result obtained when training and testing 6 representative SVM classifiers on the clinical database. Each of the SVM classifiers use a different combination of signals. Table V provides the per group classification accuracy for EXP-1. Best results on the test set for EXP-1 and EXP-2 were obtained using a combination of audio, actigraphy and ODI, achieving a test accuracy accuracy (Ac) of $Ac = 88.4\%$ and $Ac = 92.3\%$ respectively.

B. Application framework

1) *General description*: SleepAp is composed of five modules as shown in Fig. 3 which shows sample screens of the app: 1: "About apnoea" includes information about the disease, its manifestation and treatments. This module also allows the patient to listen to healthy, snoring and OSA breathing sounds;

TABLE III: EXP-1, SVM classification on clinical data: {Normal, Snorer} vs {Mild, Moderate and Severe OSA}

	Training set (735)				Test set (121)			
	Ac	Se	Sp	AUROC	Ac	Se	Sp	AUROC
ODI	84.4	80.1	88.6	0.918	85.1	80.5	93.2	0.904
AU	76.6	69.5	83.7	0.838	74.4	70.1	81.8	0.831
AC	63.1	60.5	65.8	0.673	72.7	74.0	70.5	0.792
DE	65.6	64.6	66.6	0.711	67.8	63.6	75.0	0.757
AU+AC	74.6	65.9	83.2	0.821	76.9	66.2	95.5	0.859
AU+AC+ODI	85.6	81.5	89.7	0.931	88.4	81.8	100	0.942

ODI: oxygen desaturation index, AU: audio, AC: actigraphy, DE: demographics, Ac: accuracy, Se: sensitivity, Sp: specificity, AUROC: area under receiver operating characteristic curve. Statistics are given in percentage with the exception of the AUROC.

TABLE IV: SVM classification on clinical data: EXP-2, {Normal, Snorer} vs {Moderate and Severe OSA}

	Training set (646)				Test set (104)			
	Ac	Se	Sp	AUROC	Ac	Se	Sp	AUROC
ODI	90.9	87.8	93.2	0.966	90.4	83.3	100.0	0.905
AU	81.7	68.0	92.1	0.896	77.9	66.7	93.2	0.841
AC	68.0	44.6	85.6	0.719	72.1	53.3	97.7	0.776
DE	70.7	63.7	76.1	0.744	73.1	61.7	88.6	0.764
AU+AC	81.3	71.9	88.3	0.884	79.8	70.0	93.2	0.836
AU+AC+ODI	91.3	88.1	93.8	0.978	92.3	86.7	100.0	0.942

2: “Questionnaire” is an adapted version of the clinically-validated STOP-BANG questionnaire¹ [33] which provides individuals with an initial screening tool for assessing their chance of having OSA; 3: “Record” guides the user through the setup of the microphone, armband, and pulse oximeter and records the signals of interest (audio, body movement, body position and PPG). The recorded signals and the answers from the questionnaire are stored on the phone; 4: “Analyse” applies signal processing algorithms to extract features from the signals and outputs a probability of the patient having OSA or not. The outcome can also be broken down to give the ODI and STOP-BANG questionnaire score; 5: “View previous” displays the output of previous recordings, allowing for treatment monitoring.

¹Although the STOP-BANG questionnaire is the standard questionnaire used in medical practice for OSA screening, it suffers from low specificity in spite of its high sensitivity. Additionally, the questionnaire was evaluated on a specific population (surgical patients).

2) “Record” module: Since SleepAp will be used by individuals with no background in physiological recording a set-up tutorial is provided. The images used in the tutorial are shown in Fig. 4. The user goes through the following steps; Step 1: place the phone in the armband on your preferred arm. Make sure the phone is facing directly forward as shown in the picture. Step 2: use your headphone microphone and place it next but not under the nose as shown in the picture. You can attach it with surgical tape or any other medical adhesive. Pull headphone cable up through your pyjama sleeve before taping next to your nose. Make sure that the microphone hole is not directed toward your skin. Step 3: attach the pulse oximeter on your preferred wrist. Place your finger (index, middle or ring) in the finger cuff as shown in the picture. Make sure that the cable is on the back of your hand. Step 5: some recommendations before starting:

- It is preferable to remove any third party sources of noise such as TV, radio, opened window. It is also best to sleep in a room on your own.
- If you manually stop the recording at any time before the app has finished, you will need to perform the test again.
- The recording is going to start after you press ‘Start’.

SleepAp applies the signal processing algorithm to the physiological signal recorded 30 minutes after the ‘Start’ button is pressed and for a period of four hours. This is to allow the user time to fall asleep. This time lapse is accounted for when deriving features from the signals.

C. Signals & Machine learning

Average time for MSE calculation was $9.8s \pm 1.0$ for audio and $12.2s \pm 1.0$ for actigraphy. Tests were performed with Google Nexus 7 device on 20 clinical recordings of audio and actigraphy, each lasting 4 hours. The entire process of loading recordings as well as performing MSE calculations and SVM classification was completed in $13.2s \pm 1.2$ when computing MSE in parallel processing for audio and actigraphy.

IV. DISCUSSION AND CONCLUSION

The SVM performance was reported with and without the mild OSA patients. EXP-2 (i.e. excluding mild OSA patients

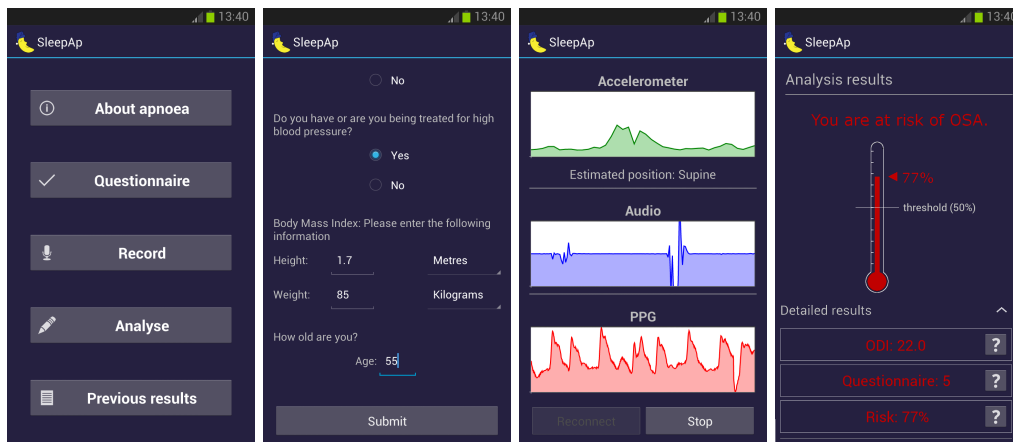


Fig. 3: SleepAp screenshots. From left to right: main menu, STOP-BANG questionnaire, recording and analysis modules.

TABLE V: Per group classification accuracy (%) for EXP-1 using various sensor configurations

	Nb patients		ODI		AU		AC		DE		AU+AC		AU+AC+ODI	
	Train	Test	Train*	Test*	Train*	Test*	Train*	Test*	Train*	Test*	Train*	Test*	Train*	Test*
Normal	130	25	96.2	96.0	86.9	76.0	65.4	64.0	71.5	76.0	85.4	92.0	97.7	100.0
Snoring	238	19	84.5	89.5	81.9	89.5	66.0	78.9	63.9	73.7	81.9	100.0	85.3	100.0
Mild	89	17	40.4	47.1	37.1	52.9	43.8	64.7	49.4	52.9	30.3	35.3	41.6	58.8
Moderate	101	22	87.1	90.9	72.3	54.5	53.5	63.6	62.4	59.1	64.4	59.1	90.1	90.9
Severe	177	38	96.0	89.5	84.2	86.8	72.9	84.2	73.4	71.1	84.7	84.2	97.7	89.5

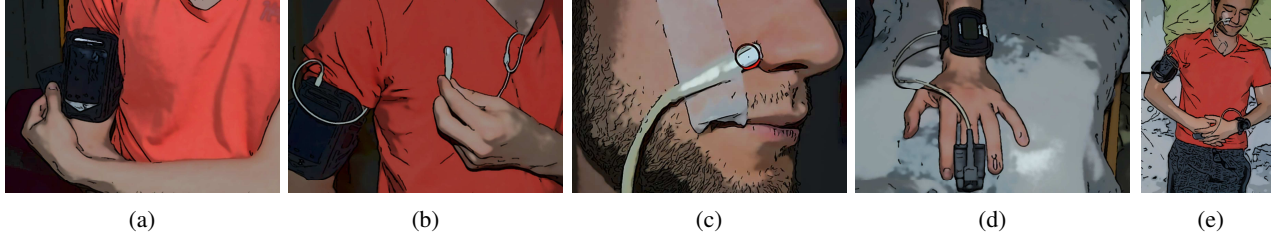


Fig. 4: Setup of the equipment; (a) attaching armband, (b) and (c) microphone setup, (d) pulse oximeter setup, (e) final setup.

in the training and test sets) was performed since in the context of SleepAp the consequences associated with wrongly classifying a mild OSA patient are minimal compared to the consequences of referring healthy users to the sleep clinic (which would dramatically increase costs) or not referring OSA users with moderate to severe OSA (which could have adverse health consequences). Classifying mild OSA was the most difficult task and had the lowest accuracy. When using the SVM structure trained in EXP-2 (i.e. without the mild-OSA patients) on all mild-OSA patients, 3/4 of the patients from this group were classified as non-OSA.

According to Table V actigraphy and audio did not provide any extra useful information for moderate OSA ($Ac \sim 50\%$ on the test sets), but they did for all other classes.

The system was designed for two types of users: 1) individuals who want to screen themselves because they believe that they are at risk of OSA, 2) medical professionals who want to use the system in the same manner as a home sleep recording system in order to diagnose patients with OSA.

The framework was designed to be modular in that classification can be obtained from any subset of the signals that are recorded. This is particularly important for two reasons: i) if the microphone, pulse oximeter or armband become dislodged overnight then the analysis can still be performed on the signals that have been correctly recorded; ii) considering the first type of user, many individuals might not have an armband or pulse oximeter readily available. For these users, the app should be a low cost system using the sensors already available in smartphones, and it should be easy to use (since the user is likely to have no training in physiological monitoring). If a pulse oximeter were essential, the proposed sleep monitoring system would dramatically increase in price and accessibility would be reduced (except if a low cost pulse oximeter that can be connected to the phone were to be released), particularly for users in developing countries.

There is also room to improve the framework by integrating signal quality indices (SQI), allowing flagging or removal of noisy sections of the PPG that are likely to affect the ODI computation, as in [34].

It was important to consider that Android does not ensure high temporal precision on the acquisition of sensor data. Thus, to increase the sampling rate precision of the accelerometer, the acquisition rate was set to 10 times greater than the desired one (i.e. 40 Hz for a 4 Hz recording) and then down-sampled by choosing samples with time stamps closer to the desired 4 Hz rate. This extra step decreased the mean absolute error from the desired sampling rate (4 Hz) from 0.99 Hz to 0.35 Hz in tests performed with a Samsung Galaxy S3 device.

Recording the audio signal on the phone requires a microphone, the location of which is important to record good quality sound, particularly if an average headset microphone is used (as opposed to the high-end quality microphones used in sleep clinics). The natural idea of placing the microphone under the nose is not actually optimal for this type of analysis; the noise induced by airflow in this location was found to dominate over the respiration pattern, compromising the analysis of the frequency content. On the other hand, placing the microphone too far from the nose (e.g. on the clothing or next to the ear) fails to acquire quieter breathing sounds. For these reasons, 'next to but not under the nose' was identified as a good compromise between sound amplitude and noise.

Although body position is recorded overnight using SleepAp, it was not possible to add it as a feature to the SVM since this feature was not recorded in the clinical database that was used to train the classifiers. This is also true of some of the answers in the STOP-BANG questionnaire that were not available for the clinical database and as a consequence were not considered as features to train the classifier. Inclusion of this additional information could further improve classification.

The sample size used to develop this application was relatively small, all patients were referred for screening by their general practitioner and most importantly the data used was not recorded using smartphones. As such, this does not represent the typical user and recording setting so retraining of the classifiers is likely to be required before any such system is used in clinical practice.

The mobile phone app presented in this paper is a first

step towards a clinically-validated automated sleep screening system which is available on a mass scale and at negligible cost to smartphone users. The general approach of selecting features which can be recorded from smartphones and training a machine learning algorithm to screen populations is likely to be of tremendous help in low resource countries with little access to expensive screening procedures.

To our knowledge, this is the first fully-automated smartphone medical diagnostic system that uses more than simple logic branching and decision support. Novel signal processing techniques together with a machine learning approach provide a simple home-use system which can output a probability of a user requiring further hospital screening for OSA. **Such software provides a new, easy-to-use, low-cost and widely available modality for OSA screening for smartphone users for the cost of a pulse oximeter (which can be as low as \$20) or at negligible cost (when no oximeter is used).** Moreover, it provides a generic framework for health screening in low resource environments. We are releasing the code under an open source license so as to encourage other users to test it on sub-populations and continue to develop the platform. **The source code is available at <http://code.google.com/p/sleep-ap/>.**

ACKNOWLEDGMENT

JB is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and the Balliol French Anderson scholarship fund. AR, JD, NP, AH are supported by the RCUK Digital Economy Programme grant number EP/G036861/1. AH acknowledges Capes Foundation, Ministry of Education of Brazil, Brasilia - DF 7004-020, Brazil (Proc. No BEX 0725/12-9). GC acknowledges the support of the Wellcome Trust through Centre Grant No. 098461/Z/12/Z for The University of Oxford Sleep and Circadian Neuroscience Institute (SCNi) and EPSRC grant EP/K020161/1. This article is dedicated to the memory of Lyn Davies who provided excellence in insight, equipment and dedication to the field.

REFERENCES

- [1] L. J. Epstein and G. R. Dorlac, "Cost-effectiveness analysis of nocturnal oximetry as a method of screening for sleep apnea-hypopnea syndrome," *CHEST Journal*, vol. 113, no. 1, pp. 97–103, 1998.
- [2] M. Vishesh Kapur, R. E. Sandblom, R. Hert, B. James, and D. Sean, "The medical cost of undiagnosed sleep apnea," *Sleep*, vol. 22, no. 6, p. 749, 1999.
- [3] H. Bearpark, L. Elliott, R. Grunstein, S. Cullen, H. Schneider, W. Althaus, and C. Sullivan, "Snoring and sleep apnea. A population study in Australian men," *Am J Resp Crit Care*, vol. 151, no. 5, pp. 1459–65, 1995.
- [4] E. O. Bixler, A. N. Vgontzas, H.-M. Lin, T. Ten Have, J. Rein, A. Vela-Bueno, and A. Kales, "Prevalence of sleep-disordered breathing in women. Effects of gender," *Am J Resp Crit Care*, vol. 163, no. 3, pp. 608–13, 2001.
- [5] M. S. Ip, B. Lam, I. J. Lauder, K. W. Tsang, K. F. Chung, Y. W. Mok, and W. K. Lam, "A community study of sleep-disordered breathing in middle-aged Chinese men in Hong Kong," *Chest*, vol. 119, no. 1, pp. 62–9, 2001.
- [6] B. Lam, D. Lam, and M. Ip, "Obstructive sleep apnoea in Asia," *Int J Tuberc Lung D*, vol. 11, no. 1, pp. 2–11, 2007.
- [7] Z. F. Udwadia, A. V. Doshi, S. G. Lonkar, and C. I. Singh, "Prevalence of sleep-disordered breathing and sleep apnea in middle-aged urban Indian men," *Am J Resp Crit Care*, vol. 169, no. 2, pp. 168–73, 2004.
- [8] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New Engl J Med*, vol. 328, pp. 1230–5, Apr 1993.
- [9] S. K. Sharma, S. Kumpawat, A. Banga, and A. Goel, "Prevalence and risk factors of obstructive sleep apnea syndrome in a population of Delhi, India," *Chest*, vol. 130, no. 1, pp. 149–56, 2006.
- [10] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705–6, 1997.
- [11] J. Behar, A. Roebuck, J. S. Domingos, E. Geder, and G. D. Clifford, "A review of current sleep screening applications for smartphones," *Phy Meas*, vol. 34, no. 7, p. R29, 2013.
- [12] A. Roebuck and G. D. Clifford, *A comparative analysis of polysomnographic signals for classifying Obstructive Sleep Apnoea*. PhD thesis, University of Oxford, 2013.
- [13] S. E. Craig, M. Kohler, D. Nicoll, D. J. Bratton, A. Nunn, R. Davies, and J. Stradling, "Continuous positive airway pressure improves sleepiness but not calculated vascular risk in patients with minimally symptomatic obstructive sleep apnoea: the mosaic randomised controlled trial," *Thorax*, vol. 67, no. 12, pp. 1090–6, 2012.
- [14] A. Roebuck, V. Monasterio, E. Geder, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. D. Clifford, "A review of signals used in sleep analysis," *Phys Meas*, vol. 35, no. 1, p. R1, 2014.
- [15] A. Roebuck and G. D. Clifford, "Multiscale entropy applied to audio data for classifying obstructive sleep apnoea patients," in *Am Thor Soc 2012 Conf.*, May 2012.
- [16] J. Behar supervised by G.D. Clifford, "Analysis of accelerometer data for apnea screening," *MSc. Thesis, University of Oxford.*, August 2011.
- [17] N. Higgins and G. Clifford, "The detection of obstructive sleep apnoea using a mobile phone," *MEng. Thesis, University of Oxford*, May 2012.
- [18] M. Costa, "Multiscale entropy analysis." <http://physionet.org/physiotools/mse/>, August 2004.
- [19] C. Hang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intel. Sys. and Tech.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [20] C. Chang and C. Lin, "Libsvm – a library for support vector machines." <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [21] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 1. Springer New York, 2006.
- [22] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, 2004.
- [23] Android, Platform Versions, "http://developer.android.com/about/dashboards/index.html," Consulted July 2013.
- [24] Canalys, "Google's Android becomes the world's leading smart phone platform." Available from: <http://www.canalys.com>, January 2011.
- [25] T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, T. Brown, A. Chesson, J. Coleman, T. Lee-Chiong, J. Pancer, and T. J. Swick, "Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007," *Sleep*, vol. 30, pp. 519–29, May 2007.
- [26] A. Oksenberg, I. Khamaysi, D. S. Silverberg, and A. Tarasiuk, "Association of body position with severity of apneic events in patients with severe nonpositional obstructive sleep apnea," *Chest*, vol. 118, no. 4, pp. 1018–24, 2000.
- [27] B. Kemp, A. J. Janssen, and B. van der Kamp, "Body position can be monitored in 3D using miniature accelerometers and earth-magnetic field sensors," *Electroen Clin Neuro*, vol. 109, pp. 484–8, Dec. 1998.
- [28] J. Solà-Soler, J. A. Fiz, J. Morera, and R. Jané, "Multiclass classification of subjects with sleep apnoea-hypopnoea syndrome through snoring analysis," *Med Eng Phys*, vol. 34, pp. 1213–20, Dec. 2012.
- [29] N. Ben-Israel, A. Tarasiuk, and Y. Zigel, "Nocturnal sound analysis for the diagnosis of obstructive sleep apnea," *Annual Int. Conf. IEEE Eng. Med. and Bio. Soc.*, vol. 2010, pp. 6146–9, Jan. 2010.
- [30] E. Dafna, A. Tarasiuk, and Y. Zigel, "Sleep-quality assessment from full night audio recordings of sleep apnea patients," *Int. Conf. IEEE Eng. Med. Bio. Soc.*, vol. 2012, pp. 3660–3, Jan. 2012.
- [31] H. Xu, W. Huang, L. Yu, and L. Chen, "Sound spectral analysis of snoring sound and site of obstruction in obstructive sleep apnea syndrome," *Acta oto-laryngologica*, vol. 130, pp. 1175–9, Oct. 2010.
- [32] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans Int Sys Tech*, vol. 2, pp. 27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [33] F. Chung, B. Yegneswaran, P. Liao, S. Chung, S. Vairavanathan, S. Islam, A. Khajehdehi, and C. Shapiro, "Stop questionnaire: A tool to screen patients for obstructive sleep apnea," *Anesthesiol*, vol. 108, pp. 812–21, 2008.
- [34] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Phy. Meas.*, vol. 33, no. 9, p. 1491, 2012.