Sleep Apnoea: a Serious Public Health Concern with Tough Information Processing Challenges

Udantha R. Abeyratne, A.S. Wakwella and A.S. Karunajeewa
*School of Information Technology and Electrical Engineering
The University of Queensland, St Lucia, Brisbane, QLD4072, Australia
Email: udantha@itee.uq.edu.au

Abstract—Obstructive Sleep Apnoea (OSA) is a serious and widespread disease caused by the collapse of the upper air passage during sleep. The gold standard of OSA diagnosis, Polysomnography (PSG) requires an overnight hospital stay, connected to over 15 channels of measurements requiring physical contact with sensors. PSG is expensive, inconvenient and is not suitable for mass screening of the population. In this paper, we discuss the medical consequences and economic impact of OSA. We also propose a novel low-cost technology based on non-contact instrumentation to diagnose the disease and evaluate its performance using clinical data.

I. INTRODUCTION

Obstructive sleep apnea is one of the most common sleep disorders. It is characterized by repetitive obstruction of the upper airways during sleep. The frequency of such events can range up to hundreds of events per sleep-hour. Full closure of the airways is termed Apnea and a partial closure is known as Hypopnea. The number of Apnea/hypopnea events per hour is known as the AHI-index, and is used by clinical community as a measure of the severity of OSA.

OSA, when untreated, presents as a major public health concern throughout the world. OSA patients use health facilities at twice the average rate [1], causing huge pressures on national healthcare systems. OSA is associated with serious complications such as cardiovascular disease, stroke, [2]-[4] and sexual impotence. It also causes cognitive deficiencies, low IQ in children, fatigue and accidents [2],[3]. Australian Sleep Association reported [4] that in the state of New South Wales alone 11,000–43,000 traffic accidents per year were attributable to untreated-OSA.

OSA is a highly prevalent disease in the society. An estimated 9% of the women and 24% of the men in the US population of 30 to 60 years was found to be having at least mild OSA [5]. In Singapore, about 15% of the total population has been estimated to be at risk [6]. In a recent study in India [7], 19.5% of people coming for routine health checks were found to have at least mild OSA. It should be noted, however, most of the prevalence studies existing today have been conducted on white Caucasians, and comprehensive data on ethnic minorities, Asians and Africans remain to be gathered.

The full clinical significance of OSA has only recently been understood. Partly as a result of this, the public awareness of the disease is severely lacking. Healthcare systems around the world are largely unprepared to cater to the massive number of OSA patients. This problem is
especially severe in the developing world, where OSA diagnostic facilities are rare to find.

A. The Standard Method of Diagnosis

The current standard of diagnosis is Polysomnography (PSG). Routine PSG requires that the patients sleep for a night in a hospital Sleep Laboratory, under video observation. In a typical PSG session, signals/parameters such as ECG, EEG, EMG, EOG, nasal/oral airflow, respiratory effort, body positions, body movements and the blood oxygen saturation are carefully monitored (see Fig.1). Altogether, a PSG test involves over 15 channels of measurements requiring physical contact with the patient.

B. Drawbacks of PSG and Possible Improvements

The routine hospital-based PSG test is the definitive method of diagnosis of the disease. However, it suffers from the following drawbacks:

• **PSG requires contact instrumentation.** Furthermore, electrophysiological channels such as EEG/ECG/EOG/EMG require more restrictive Galvanic contact with the patient. Some patients find it difficult to fall asleep in an unfamiliar surrounding connected to a host of sensors via wires. It is especially unsuited for paediatric use.

• **Poor data integrity is a common problem** in routine PSG tests. Even when the test is done in the hospital, it is common to see cases of data loss (or quality deterioration) due to various reasons (eg: improper sensor contact due to electrodes/sensors coming loose, SpO2 sensor falling off, and, measurement problems such as movement artifacts).

• **PSG interpretation is a tedious task** due to the size of the data gathered, complexity of the signals and measurement problems such as data loss. Clinicians routinely resort to manual analysis of data via visualization ('eye-balling'); even though modern PSG systems offer automatic sleep analysis software, they have not yet become sophisticated enough to gain the full trust of clinical community.

• **PSG is not suitable for mass screening** of the population. A trained medical technician is required to connect the patient to the PSG equipment, and the patient needs to be monitored overnight to avoid incurring data losses.

• **PSG is expensive;** this is another factor working against mass screening uses.

There is an enormous clinical need for a simplified diagnostic instrument capable of convenient and reliable diagnosis/screening of OSA at a home setting [8].

There has been a flurry of recent activities at developing technology to address this need. Four different classes of OSA monitors are under development [8 and references therein]. These devices vary from two channel (eg: airflow and oximetry) systems (designated Type-IV) to miniaturized full-PSG (Type-I) units [8]. Their major drawbacks are:

- existing take-home devices have at least one sensor which requires physical contact. This makes them difficult to use by untrained persons, and cumbersome to use on paediatric populations. TYPE-IV systems, with the smallest number of sensors, suffers from the fact that oximetry identifies oxygen saturation in blood only as a surrogate for OSA, and the absence of significant desaturation does not mean the absence of the disease [8],[9].

• Presence of a medical technologist is still required at the site of the test, if acceptable sensitivity/specificity performance is required. High rates of data loss (up to 20% loss at home compared with 5% at sleep lab) [8], [9] have been reported when a medical technologist is not in attendance. Unattended systems have not led to high enough sensitivity/specificity levels to be used in a routine home monitoring or a community screening exercise [8].

• Type-I and II devices use channel counts from 7-18 and are difficult to use by an untrained person. The quality and the loss of data is a serious problem.

In this paper we address these issues and present a novel instrumentation and signal processing framework suitable for developing a mass screening device for OSA. The proposed method uses non-contact measurements and thereby bypasses some of the most serious problems faced by current state-of-the-art devices.

C. Snoring and OSA – The Development of the Concept

Snoring almost always accompanies OSA and is universally recognized as its earliest symptom [2],[6],[10]. Logic dictates that it should be providing us with the earliest opportunity to diagnose the disease. At present, however, quantitative analysis of snore sounds is not a practice in OSA detection [10]. The vast potential of snoring in the non-invasive diagnosis of OSA remains unused.

In this paper, we argue that the human speech and SRS share many similarities (see Fig.2), and biological ‘wetware’ used for the generation processes. The upper airway acts as an acoustic filter during the production of snoring sounds, just as the vocal tract does in the case of speech sounds. Episodes of OSA, by definition, are associated with partial or full collapse of upper airways. As
such, changes to the upper airways brought about by this collapse should be embedded in snoring signals.

In Section II we describe how we propose to use snore sounds in detecting OSA.

II. METHODS

A. Snore/PSG Data Acquisition

The work reported in this paper depends on snore/PSG data acquired from patients undergoing routine PSG testing at the sleep diagnostic laboratory of the Princess Alexandra Hospital, Australia.

Routine PSG data were obtained from clinical equipment (Siesta, Compumedics®, Sydney, Australia). PSG data are accompanied by an edited annotation file with an event-by-event description of Apneas. A detailed sleep summary compiled during the expert scoring of sleep events at the clinic was also available.

Snore sounds were acquired in synchrony with routine PSG, using the high fidelity, CD-quality computerized data acquisition system shown in Fig.3. A matched pair of low noise microphones having a hypercardioid beam pattern (Model NT3, RODE®, Sydney, Australia) was used to capture the sound signals. The nominal distance from the microphone to the mouth of the patient was 50cm. An A/D converter unit (Model Mobile-Pre USB, M-Audio®, California, USA) was used for sound signal acquisition, at a sampling rate of 44.1k samples/s and a bit resolution of 16bits/sample. The system had a 3dB bandwidth of 20.8kHz. The size of the patient database used is 45 subjects.

B. Snore Sounds: A Working Definition

A typical PSG test runs for up to 8 hours, and it is quite common to observe up to 8,000 events of snorers within a recorded sound data. Manual segmentation of snore events is thus not feasible in mass screening applications.

One of the major problems towards automation is that there is no objective definition of what a ‘snore’ is [10]. Recently we proposed [11] an objective definition for ‘snoring’ independent of the sound intensity. It is based on the observation that sounds perceived as ‘snores’ by humans are characterized by repetitive packets of energy that are responsible for creating the vibratory sound peculiar to snorers (Fig.4: bottom).

We call the distance between such packets as the ‘pitch’ of snoring. A snoring episode (SE) will consist of a segment with pitch (‘voiced-segment’) , and possibly ‘silence’ and segments without pitch (‘unvoiced segment’). An inspiration-expiration cycle without any segment associated with a pitch is termed as a pure breathing episode (PB).

C. A Mathematical Framework for Sound Analysis

Snoring originates from acoustical energy released by the vibratory motions of upper airway site(s) during sleep. The airway cavity acts as an acoustical filter, and modifies the source sounds to produce snoring sounds we hear. Inspired by the source/vocal-tract system for human speech synthesis, we model a block of recorded sound s(n) as a convolution between: (i) ‘source signal x(n)’ representing the source acoustical energy and, (ii) ‘(TAR) Total Airway Response h(n)’ which captures the acoustical features of the upper airways as given by:

\[ s(n) = h(n) * x(n) + b(n) \]  

The symbol b(n) denotes background activities, x(n) is the source excitation, “*” denotes convolution and h(n) is the TAR function. The term b(n) is due to a range of independent reasons and hence considered a Gaussian distribution independent of x(n).

The nature of x(n) depends on the nature of the sound segment under consideration. For the case of a segment without any pitch (i.e., either a PB or unvoiced-snoring segment), we model the source excitation, x(n) as a white random process. The x(n) for voiced snoring segments is modeled by a pseudo-periodic pulse–train, drawing from techniques used in speech analysis.

The TAR function h(n) is modeled in this paper as a mixed-phase system considering that in OSA, multiple-source locations with temporal relations between each other can be found and thus phase-information cannot be neglected in general.
It is our hypothesis that the state of the upper airways can be characterized by the pair \( \zeta = \{ g[x(n)], f[h(n)] \} \), where \( g \) an \( f \) represent the operation of extracting features out of the \( x(n) \) and \( h(n) \).

D. Sound Segmentation

We pioneered an automated snore segmentation algorithm [11] and implemented it as a Matlab \(^b \) based software package. The package can take in full night sound data and separate snoring and pure-breathing episodes. It then further divides snore episodes into sub-categories ‘voiced’, ‘unvoiced’ and ‘silence’ sections. In this paper, we use the segmentation algorithm developed in [11] as a pre-processing stage of our technique. The segmentation technique proved to be 99.6% accurate.

E. Estimating the pitch ‘p’ of snoring (p= g[x(n)])

The segments of sound recording that has been determined to have a pitch (i.e. segments with internal periodicity) have been further analyzed to estimate the pitch associated with each segment.

Consider an arbitrary j-th Snoring Episode (SE) in the sound recordings. Divide the voiced segment \( s_{\nu,j} \) of the j-th Snoring Episode into \( L_j \) number of data blocks \( \{ B_k \} \), \( k = 1... L_j \), each of length \( N \). Thus, at the output of the pitch-detector, each data block in the set \( \{ B_k \} \) is associated with a pitch period \( \mu_k \). We term the series \( \{ \mu_k \} \), \( k=1,2,...,L_j \) as the Intra-snore Pitch Series for the j-th Snoring Episode.

We consider the structure of the Intra-snore Pitch Series, and show that it is characteristic by discontinuities, which can be used in the diagnosis of OSA. We propose a new measure called ISPJ-Probability to capture intra-snore jumps in pitch, via Definition-1 and Definition-2 given below:

**Definition-I:** Suppose that in the j-th arbitrary snoring episode \( s_{\nu,j} \) there are at least \( q \) (\( < N_j +1 \)) data blocks in the set \( \{ B_k \} \), \( k=1... N_j \) with pitch periods \( \mu_k \) greater than a pitch threshold \( \gamma \). Then the entire Snoring Episode \( j \) is labeled as having the feature ISPJ at level \( q, \gamma \). We call the quantity \( q \gamma \) as the ‘jump multiplicity’.

**Definition II:** Define a quantity ISPJ-probability \( P_{q}(\tau_{r_0}) \) at level \( q, \gamma \) for the signal \( s(n) \) of a length \( D \) as: \( P_{q}(\tau_{r_0}) = 100 \frac{n_{q}(\tau_{r_0})}{n_{0}(r_0)} \)%, \( n_{0}(r_0) \) is the total number of snore episodes contained within the data of length \( D \). The quantity \( n_{q}(\tau_{r_0}) \) is the number of episodes within \( D \) that were labeled as having the feature ISPJ according to Definition-I.

The sound source based test statistic proposed in this paper is \( P_{q}(\tau_{r_0}) \), with the corresponding decision threshold symbolized by \( P_{th} \). If \( P_{q}(\tau_{r_0}) > P_{th} \), the snore-based test is positive for OSA, and vice versa.

For each subject in the database, a set \( S \) of ISPJ-Probabilities is defined by:

\[
S = \{ P_{q}(\tau_{r_0}) \ | \ q = 1, 2, 3 \ and \ \gamma = 10, 11, ...25 \text{ms} \}, \tag{3}
\]

where \( P_{q}(\tau_{r_0}) \) is calculated via (2). We use \( S \) to derive ROC curves of \( P_{q}(\tau_{r_0}) \) at different ISPJ levels \( q, \gamma \).

To draw Receiver Operating Characteristics (ROC curves), we need to know the “true clinical diagnosis” of the patients. The PSG based diagnosis is considered the absolute truth in this paper. The diagnosis based on the test statistic \( P_{q}(\tau_{r_0}) > P_{th} \) is compared to the ‘absolute truth’, and the nature of the decision is noted as one among (i) true positives (TP), (ii) true negatives (TN), (iii) false positives (FP) or (iv) false negatives (FN). Sensitivity, a measure of success in detecting TPs is defined as: \( TP/(TP + FN)\% \). Specificity, a measure of success in rejecting non-diseased subjects, is defined as \( TN/(TN + FP)\% \). The ROC curve is a graph of sensitivity vs. (1-specificity).

F. Estimating the TAR function as a Mixed-Phase Signal

The TAR (h(n)) of snoring characterizes the acoustical features of the airways, just like the vocal-tract response does in Speech. Since the disease of OSA directly affect the acoustical features of the airways, it is indeed reasonable to expect that TAR will be extremely useful to detect OSA.

We focus our attention to techniques that preserve Fourier-phase information while estimating \( h(n) \). The reason is that our preliminary studies [12] have revealed that the TAR needs phase information for a complete description. This also agrees with physiological evidence (spatiotemporal diversity) of the snore generation process [10], which leads to non-minimum phase characteristics. Conventional 2nd order statistics based methods (e.g. Autocorrelation, Power Spectrum), therefore, will not be able to estimate the TAR keeping the full information content [13],[14].

In this paper, we use HOS techniques to estimate the TAR, preserving true phase information. The third order \( c_{3}(\tau, \rho) \) of the (zero-mean) segment \( s(n) \) under study is defined as [14]:

\[
c_{3}(\tau, \rho) = E \{ s_{j}(n) s_{j}(n+\tau) s_{j}(n+\rho) \}, \tag{6}
\]

where \( E[\cdot] \) denotes the expectation operator.

The bispectrum \( C_{3}(\omega_1, \omega_2) \) [14] is defined to be the Fourier transform of the third order cumulant \( c_{3}(\tau, \rho) \). Converting (6) to the bispectrum domain we get:

\[
C_{3}(\omega_1, \omega_2) = C_{3}(\omega_1, \omega_2) H(\omega_1) H(\omega_2) H(\omega_1+\omega_2) + C_{3}(\omega_1, \omega_2) \tag{7}
\]

where \( C_{3}(\omega_1, \omega_2) \) is the bispectrum of \( x(n) \), \( H(\omega) \) is the spectrum of \( h_{th}(n) \), “*” denotes conjugate and \( C_{3}(\omega_1, \omega_2) \) is the noise bispectrum. Due to the Gaussianity of \( b(n) \), the term \( C_{3}(\omega_1, \omega_2) = 0 \) as dictated by the properties of HOS. Thus from (7) we see how HOS allows us to transform the signal to a higher Signal-to-Noise Ratio (SNR) domain for further analysis.

The bicepsrum [14] is defined as the 2-dimensional cepstrum of the bispectrum. In the bicepsrum domain, from (7) we obtain:

\[
b_{3}(m_1, m_2) = b_{3}(m_1, m_2) + b_{n}(m_1, m_2), \tag{8}
\]
where $b_s(m_1,m_2)$, $b_u(m_1,m_2)$ and $b_b(m_1,m_2)$ are the biceps of $s(n)$, $h(n)$ and $s(n)$ respectively. Note that the noise term $b(n)$ disappears in (4), because HOS is theoretically immune to Gaussian additive noise [14].

Now the problems at hand are to estimate (i) the $\text{TAR}$ ($h(n)$), and (ii) the properties of the source excitation $x(n)$ as relevant to the voiced and unvoiced snore segment.

As well known in the HOS signal processing community the slice $b_s(m,0)$ of $b_b(m_1,m_2)$ gives the 1-D complex cepstrum of $s(n)$, which can be used to reconstruct the signal $s(n)$ or extract $x(n)$ and $h(n)$ through inverse 1-D complex cepstra operations.

Thus, the methods proposed so far allows us to estimate $\zeta = \{P\gamma, h(n)\}$ as a method to characterize the upper airways.

III. RESULTS

A. Pitch and the ISPJ-Probability

In Fig. 5 we illustrate a typical result we obtained from our pitch calculation operations. Fig.5(a) shows a section of the sound recording with four breathing episodes. Fig. 5(b) shows the outcome of our segmentation algorithm. Note that all episodes shown in the figure qualify as snoring episodes; within each episode we see segments with well defined periodicity (‘voiced segments’).

In Fig.5(c) we illustrate the concept, ISPJ proposed in this paper. The symbol ‘*’ in Fig.5(c) represents pitch values computed from each block $\{B_k\}$ $k=1...N$, of the voiced segments $\{s_{vo}\}$ $j=1,2,3,4$. Intra-snore pitch discontinuities (jumps) are clearly visible in Fig. 5(c). It is our hypothesis that these jumps, as captured in the novel measure ISPJ-Probability, will enable us to diagnose OSA based on snore sounds alone.

Next, we investigate the performance of the ISPJ-probability measure in diagnosing OSA, using our PSG/snore database.

In Fig. 6 we show ROC curves at ISPJ-levels $(2,\gamma)$ for different values of $\text{AHI}_{ib}$. These ROC curves were generated by changing $P_{\text{ib}}$ from 0.05 to 1.0 in steps of 0.05, and estimating sensitivity and specificity of detection based on the set $S$. These curves allow us to pick an optimum pitch threshold $\gamma$ which result in the best sensitivity/specificity trade-off at a desired $\text{AHI}_{ib}$.

ISPJ-levels $(1,\gamma),(3,\gamma)$ and $(4,\gamma)$ lead to results that are qualitatively similar to the ones obtained with $(2,\gamma)$, but are quantitatively inferior in their sensitivity/specificity performance.

According to Fig.6, ISPJ-Probability proposed in this paper has potential in developing a home-based Apnoea detection system targeted for community screening. It should be noted that the sensitivity/specificity measures obtained with the proposed method are on a par with the best of the comparable methods reported in the comprehensive survey of [1]. However, the unmatched advantage of the proposed method is that snore acquisition requires no sensor-contact with the subject.

B. Total Upper Airway Response

In order to investigate the behaviour of TAR we analysed one night’s SRS data from our database. In this paper we show a typical example result we obtained from a patient diagnosed with severe OSA (AHI index=100).

The patient displays pure Breathing Episode (PB) for the first 32 minutes in bed, after which he develops the first event of hypopnea at $t = 34.7$ mins. In Fig. 7(a) we show a series of TAR (breathing) estimated from a set of data blocks from PB before the patient developed the first hypopnea.

According to Fig. 7(a), all the PB before the onset of apnoea show similar characteristics. They produce slowly varying spikes with a dominant central lobe. Fig.8(a) shows the corresponding Fourier amplitude spectra of one PB, which show that the TAR is a low pass signal with most of its energy concentrated below 300Hz.

Frames (c)–(f) shown in Fig. 7 correspond to a 3-hour period during which the patient goes through a series of
fully developed OSA. The first 60ms data block of SE or PB immediately after the patient emerges from OSA events were used to estimate the TARs shown in Fig. 7(c), (d), (e) and (f). Thus they represent the acoustical state of the upper airways immediately after a full closure and opening.

According to frames (c), (d), (e) and (f) of Figs. 7 and 8, TARs of voiced snoring (immediately after OSA) are significantly different from TAR-PB, before any abnormal events. The former is richer in features and essentially register as a set of band-pass signals with a centre frequency in the range 500-1200 Hz.

It is important to find out the nature of TAR before the onset of an OSA event. In Fig. 7(b) we show a TAR (SE) estimated from the very first SE of the patient (at t = 31.2 mins) which had been classified as a benign snore. Coincidently, that particular event was followed 3 minutes later by the first hypopnoea event at t = 34.7 min. Fig. 7(b) and Fig. 8(b) both point out that the TAR for this snore is different from a TAR-PB. Rather than displaying low-pass characteristics like TAR-PB, it shows band-pass characteristics like TAR-SE, but at a much lower value of centre frequency.

IV. CONCLUSION

In this paper we explored the snore sound-based non-contact diagnosis of OSA. Drawing from speech processing literature we developed a source/TAR model for the description of upper airways during sleep, and provided new features to characterize OSA events. Our results indicate that the performance of the feature Intra-Snore-Pitch-Jump Probability (ISPJ-probability), is on a par with other competing technologies in terms of the sensitivity/specificity characteristics. However, snore based diagnosis proposed in this paper is superior in that it does not involve contact instrumentation, thus solving a major problem in population screening of the disease. The total airway response appears to capture the changes of the upper airways, but more extensive investigations are required to evaluate its performance, both in isolation and in conjunction with ISPJ-probability.

ACKNOWLEDGEMENT

The author wish to thank Dr. Craig Hukins and Ms. Lyn Fraser, Princess Alexandra Hospital, Australia for access to routine PSG data. This work was partly sponsored by a University of Queensland Early Career Researcher Grant to the first Author.

REFERENCES