AUTOMATIC CLASSIFICATION OF BREATHING SOUNDS DURING SLEEP

Brian R. Snider

Oregon Health & Science University
Center for Spoken Language Understanding
Portland, OR
sniderb@ohsu.edu

Alexander Kain

BioSpeech, Inc.
940 Upper Devon Lane
Lake Oswego, OR
kain@biospeech.com

ABSTRACT

Sleep-disordered breathing (SDB) is a highly prevalent condition associated with many adverse health problems. As the current means of diagnosis (polysomnography) is obtrusive and ill-suited for mass screening of the population, we explore a non-contact, automatic approach that uses acoustics-based methods. We present a method for automatically classifying breathing sounds produced during sleep. We compare the performance of several acoustic feature representations for detecting diagnostically-relevant sleep breathing events to predict overall SDB severity. Our subject-independent method tracks rest in the breathing cycle with 84–87% accuracy, and predicts SDB severity at a level similar to polysomnography.

Index Terms— sleep apnea, breathing, polysomnography

1. INTRODUCTION

Sleep-disordered breathing (SDB) is believed to be a widespread, under-diagnosed condition associated with many detrimental health problems [1, 2]. Young et al. describe the total burden of sleep-disordered breathing on the health system and society as "staggering" [3]. The current gold standard for diagnosis of sleep-disordered breathing is a sleep study, or polysomnography (PSG). This overnight procedure takes place in a sleep laboratory and is obtrusive, typically recording twelve or more biological processes (including electroencephalogram, electrooculogram, electromyogram, blood oxygen saturation, and nasal airflow) requiring 22–40 wires to be attached to the patient. Scoring of study results is also time-consuming and expensive, as an entire night-long study must be manually assessed by a human expert, then reviewed by a clinician to determine a diagnosis. Moreover, studies show that patients sleep differently at a hospital or clinic than at home [4]. Some at-home PSG systems exist, but these still require sensor attachments (e. g. face mask to measure airflow) and a degree of training to operate.

The complex clinical nature and high cost of PSG make the procedure ill-suited for mass screening of the population. Consequently, there is a tremendous need for an alternative method to screen for sleep-disordered breathing. Our current work investigates an acoustics-based system for tracking breathing during sleep to predict SDB severity. This system detects long rests in the breathing cycle and the overall incidence of snoring.

Our aim is to create algorithms for automatically detecting SDB-related events during sleep in a patient's home environment, with the long-term goal of creating an easy to use, mobile SDB screening device. By using only acoustic data from an ambient microphone without obtrusive wire leads, belts, masks, or taped-down sensors as used in PSG, we anticipate that capturing sleep data unobtrusively in a familiar environment will be more representative compared to using sensors in an unfamiliar clinical environment. Additionally, an automatic approach can provide objective measurements at a fraction of the time and cost of PSG.

This paper is organized as follows: after reviewing relevant literature (Section 2), we present our data collection and annotation approach (Section 3), followed by details about our experimental setup (Section 4). We then discuss algorithm performance (Section 5), and conclude with a view to the future (Section 6).

2. BACKGROUND

In a 2003 survey [5], only two of 51 papers between 1990 and 2001 utilized sleep sounds for SDB diagnosis. Since then, many studies featuring the use of acoustic data have been conducted. Several studies focus on snore detection, as snoring is seen as a possible indicator for the most common form of SDB, obstructive sleep apnea (OSA) [6, 7]. These studies explore ways to automate screening and diagnosis by analyzing sleep sounds using low-cost, non-contact, portable acoustic sensors and digital signal processing techniques, based on the hypothesis that snore signals carry relevant information about the state of the upper airways, especially the partial or full collapse thereof [8, 9, 10, 11, 12, 13, 14]. The best of these systems can detect OSA at sensitivities of 86–100% while holding specificity at 50–80%.

These studies use a variety of feature extraction techniques to generate a compact representation of the underlying acoustic signal, including: log energy, number of zero crossings, and intra-snore pitch jump probability [13]; Mel-frequency cepstral coefficients, delta coefficients, and acceleration coefficients [14]; average energy of 500 Hz sub-bands in the 0–7500 Hz frequency range [15]; F_1, F_2, and F_3 formants extracted from 14^{th}-order linear predictive coding (LPC) [16]; energy, number of zero crossings, and first LPC coefficient [17]; and average energy of 500 Hz sub-bands followed by principal component analysis [18]. Likewise, these studies use an assortment of classification machinery, including: Gaussian minimum-probability-of-error [13]; hidden Markov models [14]; classification boundary via linear regression [15]; linear decision boundary via mean and standard deviation [16]; minimum Euclidean distance from a probability density function [17]; and fuzzy C-means clustering [18].

In this paper, we focus on energy-independent acoustic features to permit classification across multiple subjects and better handle subject position differences. We use hidden Markov models for classification to model the notion of breathing state in the respiratory cycle. We also use the *absence* of breathing to predict overall SDB severity, instead of relying solely on the presence of snoring.

3. DATA

3.1. Data collection

We recorded four adult subjects with varying degrees of SDB severity during clinical polysomnography. The audio recordings were captured in parallel with typical PSG sensor data collected during the overnight study. Audio that was captured before falling asleep and after waking was excluded from the experiment. Likewise, we excluded audio that was captured after remedial measures were taken (e. g. positive air pressure was titrated or oxygen was administered), as the measures introduced additional airflow noise in the sleep environment near the subject's mouth and nose.

To capture high-quality audio, we used a highly directional Audio-Technica AT8035 microphone interfaced to a laptop computer via an audio capture device configured for a 16 kHz sample rate and 16-bit resolution. The microphone was affixed to an articulated microphone stand in the subject's room and oriented toward the subject's head when in a supine sleeping position.

3.2. Manual event labeling

For each subject, we identified four continuous regions of audio, each approximately four minutes in length. We selected the four regions from various times during the night to cover possible differences due to sleep stage or bed posture, various breathing patterns, and episodes of snoring. Additionally, we consulted the PSG data to ensure that actual apneic events were represented in the selected regions of audio.

We manually labeled each region according to the observed physiological event, as determined by listening to the region and visually inspecting the spectrogram. The possible physiological events and their corresponding symbols are: breath-in (*Bi*), breath-out (*Bo*), snore-in (*Si*), snore-out (*So*), and rest/no-signal (*N*). Figure 1 illustrates a brief excerpt of an example region and its associated manual event labels. Note that a single continuous breathing event may consist of more than one constituent part, such as a breath-in that turns into a snore-in. During labeling, we restricted a single inhalation or exhalation event to include up to three constituent parts. For example, a single inhalation may be as complex as *Bi*-*Si*-*Bi* (cf. Figure 1, 3.8–5.6 seconds), but not *Bi*-*Si*-*Bi*-*Si*. Only one of our subjects exhibited *So* events in the selected regions.

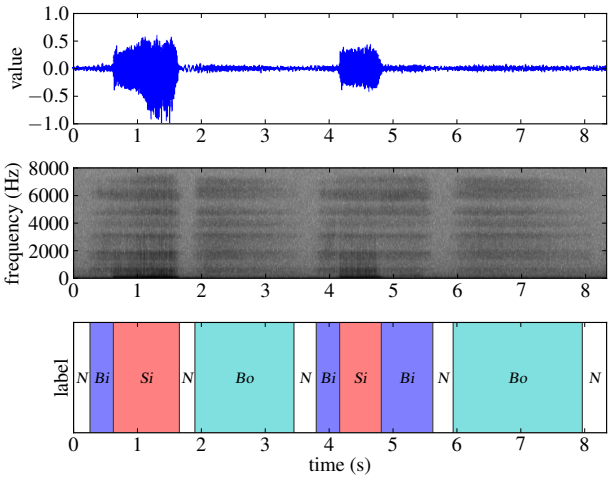

Fig. 1. Waveform, spectrogram, and manual event labels for a brief excerpt of an example region of audio.

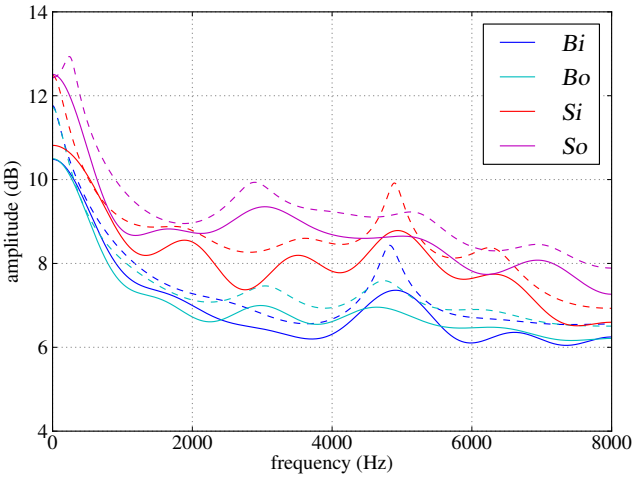

Fig. 2. Spectral reconstruction of CC (solid lines) and LPC (dashed lines) features for example events.

4. EXPERIMENT

4.1. Feature extraction

We extracted acoustic features from the audio waveform for each region using a frame length of 150 ms with zero overlap and a Hanning analysis window. We independently calculated 13 cepstral coefficients (CC), Mel-frequency cepstral coefficients (MFCC), and reflection coefficients from linear predictive coding (LPC). We excluded the first coefficient from the resulting CC and MFCC feature vectors, to make the features energy-independent. (LPC reflection coefficients already model the spectrum in an energy-independent manner.) In addition, we included first-order delta features derived from the static features.

Figure 2 illustrates the spectral reconstruction of CC and LPC features for example instances of each event type. Note that the CC features represent the spectrum smoothly, while the LPC features focus on modeling the spectral peaks. Also note that the features contain similarities across event types. For example, both *Bi* and *Si* exhibit a peak near 4800 Hz. Likewise, *Bo* and *So* both contain a peak near 2800 Hz.

4.2. Classifier

We used hidden Markov models (HMM) to predict state sequences to capitalize on the inherent notion of state in a typical respiratory cycle. We assume *a priori* that respiratory event states evident in the acoustic data can be learned and predicted by the HMM, much like phone states in speech recognition applications. Figure 3 illustrates the topology of our HMM. Note that each respiratory event type (i. e. *Bi*, *Bo*, *Si*, *So*) consist of three states per event, while the *N* event only consists of one state. We specified three states per respiratory event to capture the initial, majority, and final portions contained in a single event.

We observed many interesting respiration cycle phenomena during manual event labeling. For example, within a single inhalation, a breath-in may turn into a snore-in; likewise, during an exhalation, a snore-out may degrade into a breath-out. Additionally, we ob-

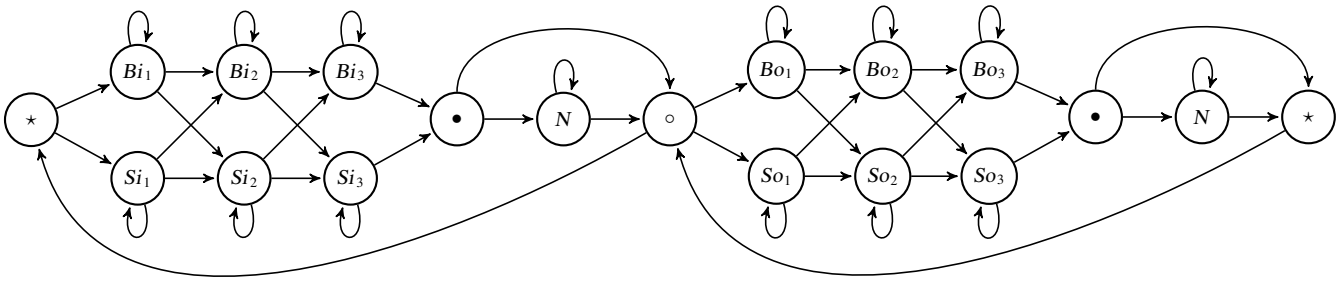

Fig. 3. HMM topology with 3 states per respiratory event type (*Bi*, *Bo*, *Si*, *So*) and 1 state per *N*. Stars ($\star$) denote the null state at the start of a respiratory cycle. Filled and open circles ($\bullet$, $\circ$) denote a null state following an in or out event. Note optional *N* event.

served that an inhalation may be immediately followed by an exhalation, with no intermediate rest (arcs emanating from filled circle states in Figure 3). Finally, we account for multiple short in or out events in rapid succession, optionally separated by rests (bottom arcs). We observed this type of phenomenon during obstructive apnea events, when a subject tried repeatedly to breathe in with limited success. We designed our model to capture these various phenomena by learning the transition probabilities between sub-event states.

4.3. Automatic label remapping

We used a simple remapping algorithm to convert the manual event labels to the state names used by our model. We assigned three states per event for inhalation and exhalation events according to the following rules: (1) if an event consists of one constituent part, divide the part into three equal-duration states; (2) if an event consists of two constituent parts, divide the longer-duration part into two equal-duration states, and assign the shorter part to the remaining state; and (3) if an event consists of three constituent parts, assign each part to a single state, preserving the original durations.

For example, an inhalation event consisting of *Bi*-*Si*-*Bi* remaps to the state sequence Bi_1, Si_2, Bi_3. Likewise, an inhalation event consisting of a brief *Bi* part that turns into a longer *Si* part remaps to the state sequence Bi_1, Si_2, Si_3, with Bi_1 having the same duration as the brief *Bi* part, and Si_2 and Si_3 equally sharing the duration of the *Si* part. Figure 4 illustrates the resulting state names for remapping events consisting of one (1.9–3.4 seconds, 5.9–7.9 seconds), two (0.3–1.7 seconds), and three (3.8–5.6 seconds) constituent parts. Note that all *N* event labels translate directly to the single *N* state.

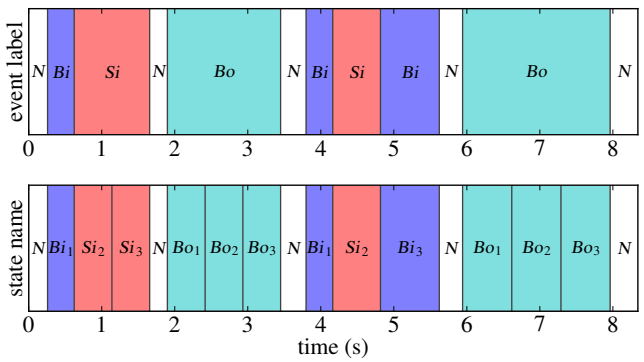

Fig. 4. Manual event labels and resulting state names after remapping for a brief excerpt of an example region of audio.

4.4. Training and testing

To train and test the system, we employed a k-fold cross-validation scheme, separating the data into different training and testing sets. We conducted both a subject-dependent (SD) and a subject-independent (SI) experiment. For the SD experiment, we considered each subject individually, creating training and testing sets by holding out one region of that subject's data. Each training set in the SD experiment was based on 12 minutes of audio, with 4 minutes held out for testing. For the SI experiment, we considered all subjects, creating training and testing sets by holding out data for one entire subject. For each fold, one subject's data was held out for testing, while the remaining three subjects' data was used for training. Each training set in the SI experiment was based on 48 minutes of audio, with 16 minutes held out for testing. In both the SD and the SI experiment, the held-out portion was cycled through all folds (k=4 regions for the SD experiment, k=4 subjects for the SI experiment), and the resulting fold's training and testing sets were used to train and test the system, respectively.

Using the state-labeled data (cf. Section 4.3), we grouped the frame-level feature vectors from the training set by state. Then, for each state, we calculated the mean and covariance of the feature vectors for that state. We then used these statistics to fit a Gaussian mixture model (GMM) with three mixture components and full covariance. Next, we initialized an HMM using the GMMs and then trained the HMM on the training set (in batch mode) using the Baum-Welch expectation-maximization algorithm with a log-likelihood change threshold of 0.01 and a maximum of 10 iterations. Finally, we used the trained HMM to decode the test set using the Viterbi search algorithm. We recorded the predicted state sequence and compared it to the manually labeled sequence.

5. RESULTS

We calculated classification performance at three levels of granularity: event-, summary-, and rest-level accuracy. For event-level accuracy, we combined states of the same type into one event (e. g. predicted states Si_1, Si_2, and Si_3 in the three-states-per-event model were all considered *Si* event frames). The event-level accuracy was used to evaluate the basic accuracy of the classifier. For summary-level accuracy, we combined in and out events of the same parent type into one category; for example, *Bi* and *Bo* frames were all considered "breath" frames. The summary-level accuracy was used to evaluate the potential for identifying snores. Finally, for rest-level accuracy, we combined all breath and snore frames into one generic "respiratory event" category, to evaluate the potential for identifying rests in the breathing cycle.

Feature Type		Accuracy		
		Event	Summary	Rest
CC				
	SD	0.85 (0.03)	0.87 (0.02)	0.90 (0.01)
	SI	0.71 (0.04)	0.73 (0.04)	0.85 (0.04)
MFCC				
	SD	0.85 (0.04)	0.86 (0.03)	0.90 (0.03)
	SI	0.73 (0.03)	0.76 (0.01)	0.84 (0.03)
LPC				
	SD	0.86 (0.04)	0.88 (0.03)	0.90 (0.03)
	SI	0.76 (0.03)	0.79 (0.04)	0.87 (0.03)

Table 1. Results for mean event-, summary-, and rest-level accuracy and standard deviation across feature types for subject-dependent (SD) and subject-independent (SI) experiments.

Table 1 summarizes our mean subject-dependent (SD) and subject-independent (SI) classification accuracy results. We found that all three feature types performed at similar levels in the SD experiment. However, in the SI experiment, LPC features outperformed other feature types at each accuracy level, making it a good candidate for use in overall system evaluation.

To evaluate our system's ability to predict overall SDB severity, we computed an apnea index (AI), a measure that indicates the number of apnea events (full airway obstruction or absence of effort) per hour. The AI is closely related to the apnea-hypopnea index (AHI), a clinical metric which includes hypopnea events (partial airway collapse) in addition to apnea events. The AHI is typically calculated during the review of many PSG signals and stratifies into none, mild, moderate, and severe SDB (AHI of $<$5, 5–15, 15–30, and $>$30, respectively) [19]. We use this same scale to translate our predicted AI and the actual AI derived from the PSG results to a corresponding SDB severity level. In the scoring criteria set forth by the American Academy of Sleep Medicine (AASM), apnea events must last at least 10 seconds [20]. In this paper, we calculated the predicted AI by using the rest-level event labels from the SI experiment with LPC features to identify all predicted *N* events with a duration of 10 seconds or greater (but less than a reasonable threshold of 60 seconds), then calculated the number of events per hour.

Additionally, we calculated the snore index, or number of snores per hour, using the summary-level event labels from the same experiment. To do so, we identified all predicted *Si* and *So* events with a minimum duration of 0.450 seconds (i. e. three frames). The snore index is not part of the clinical criteria for diagnosing SDB; however, some studies suggest that a high incidence of snoring may indicate that a subject suffers from some form of SDB [11, 12].

We computed the snore index, apnea index, and corresponding SDB severity for each subject and compared to the PSG findings. Table 2 summarizes our subject-independent results. Note that the predicted AI and SDB severity for Subject 3 are particularly inaccurate. This subject exhibited very quiet breathing out, which led to inaccurate manual event labels containing particularly long *N* events where a *Bo* event could not be discerned between two inhalation events. This discrepancy between the manual event labels and the actual respiratory events caused many false positives when searching for *N* with a duration of 10 seconds or greater. This situation could be addressed by excluding unusually long *N* portions from the AI prediction.

Subject	Predicted			Actual	
	Snore	AI	Severity	AI	Severity
1	103.8	3.7	none	0.0	none
2	346.3	0.0	none	3.2	none
3	506.1	54.6	severe	6.7	mild
4	593.0	11.1	mild	7.1	mild

Table 2. Results for predicted and actual snore index, apnea index, and SDB severity for subject-independent (SI) experiment. Actual snore index was not reported in the PSG results.

6. CONCLUSIONS AND FUTURE WORK

We presented an HMM-based classification system that predicts respiratory events during sleep from acoustic data alone. We found that the use of LPC features generally resulted in the highest accuracy at each level versus cepstral and Mel-frequency cepstral coefficient features, yielding 86–90% accuracy in the SD experiment and 76–87% accuracy in the SI experiment. We used the rest-level output from the SI experiment to predict the AI and SDB severity, with promising results. Although the precise AI differs from that derived from the clinical PSG results, a correctly predicted severity may be considered sufficient evidence that a subject should be referred to a sleep specialist. Clinical guidelines set forth by the AASM suggest that 15 or more events per hour (or 5 or more events per hour with related symptoms such as daytime sleepiness) is sufficient evidence for SDB [19]. We conclude that SDB screening may be possible by using an acoustics-based system to identify long rests in the respiration cycle to estimate the SDB severity. Furthermore, an unobtrusive acoustics-based screening system could be used at home, increasing patient comfort and capturing more representative sleep while simultaneously reducing the burden on the healthcare system to screen for SDB-related conditions.

We also used the summary-level output from the same subject-independent experiment to predict the snore index. It may be possible to use the predicted snore index to help differentiate between levels of SDB severity. The predicted snore index reported in Table 2 appears to follow the same general trend as the actual AI and SDB severity. Further analysis of the snore index is needed to determine its relation to the AI or AHI.

The accuracy of the predicted snore index, apnea index and corresponding SDB severity depends on accurate tracking of the respiratory cycle. To increase tracking accuracy, we will explore techniques (e. g. background noise reduction) to reduce confusability between similar event types, such as quiet breathing and rest events. We will also explore hybrid methods that integrate other unobtrusive sensors such as ultra-wideband radar or wireless pulse oximetry into the SDB severity prediction, to make our system more robust to respiratory cycle tracking errors.

7. ACKNOWLEDGEMENTS

This research was supported by funds from the Engineering and Technology Industry Council, Oregon Health & Science University (OHSU), and BioSpeech. OHSU and A. Kain have a significant financial interest in BioSpeech, a company that may have a commercial interest in the results of this research and technology. This potential conflict was reviewed and a management plan approved by the Conflict of Interest in Research Committee and the Integrity Program Oversight Council was implemented.

8. REFERENCES

[1] Terry Young, Paul E. Peppard, and D. J. Gottlieb, "Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective," *American Journal of Respiratory and Critical Care Medicine*, vol. 165, no. 9, pp. 1217–1239, 2002.

[2] Vishesh Kapur, Kingman P. Strohl, Susan Redline, Conrad Iber, George O Connor, and Javier Nieto, "Underdiagnosis of Sleep Apnea Syndrome in U.S. Communities," *Sleep and Breathing*, vol. 6, no. 2, pp. 49–54, 2002.

[3] Terry Young, M. Palta, J. Dempsey, Paul E. Peppard, F. J. Nieto, and K. M. Hla, "Burden of Sleep Apnea: Rationale, Design, and Major Findings of the Wisconsin Sleep Cohort Study," *WMJ : Official Publication of the State Medical Society of Wisconsin*, vol. 108, no. 5, pp. 246–249, 2009.

[4] E. Z. Osman, J. E. Osborne, P. D. Hill, and B. W. V. Lee, "Snoring assessment: do home studies and hospital studies give different results?," *Clinical Otolaryngology & Allied Sciences*, vol. 23, pp. 524–527, 1998.

[5] W. W. Flemons, "Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature: An Evidence Review Cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society," *Chest*, vol. 124, no. 4, pp. 1543–1579, 2003.

[6] V. Hoffstein, *Snoring — The Principles and Practice of Sleep Medicine*, Saunders, Philadelphia, PA, 3rd edition, 2000.

[7] Dirk Pevernagie, Ronald M. Aarts, and Micheline De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[8] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera, "Pitch analysis in snoring signals from simple snorers and patients with obstructive sleep apnea," in *Second Joint EMBS-BMES Conference 2002 24th Annual International Conference of the Engineering in Medicine and Biology Society. Annual Fall Meeting of the Biomedical Engineering Society*. 2002, pp. 1527–1528, IEEE.

[9] Udantha Ranjith Abeyratne, C. K. K. Patabandi, and Kathiravelu Puvanendran, "Pitch-jitter analysis of snoring sounds for the diagnosis of sleep apnea," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 2001, pp. 2072–2075.

[10] T. H. Lee, Udantha Ranjith Abeyratne, Kathiravelu Puvanendran, and K. L. Goh, "Formant-structure and phase-coupling analysis of human snoring sounds for the detection of obstructive sleep apnea," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 3, 2000.

[11] J. A. Fiz, "Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnea," *European Respiratory Journal*, vol. 9, pp. 2365–2370, 1996.

[12] F. Dalmasso and R. Prota, "Snoring: analysis, measurement, clinical implications and applications," *European Respiratory Journal*, vol. 9, pp. 146–159, 1996.

[13] Udantha Ranjith Abeyratne, A. S. Wakwella, and Craig Hukins, "Pitch jump probability measures for the analysis of snoring sounds in apnea," *Physiological Measurement*, vol. 26, no. 5, pp. 779–798, 2005.

[14] W. D. Duckitt, S. K. Tuomi, and T. R. Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiological Measurement*, vol. 27, no. 10, pp. 1047–1056, 2006.

[15] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, and T. Akcam, "An efficient method for snore/nonsnore classification of sleep sounds," *Physiological Measurement*, vol. 28, no. 8, pp. 841–853, 2007.

[16] Andrew Keong Ng, Tong San Koh, Eugene Baey, Teck Hock Lee, Udantha Ranjith Abeyratne, and Kathiravelu Puvanendran, "Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?," *Sleep Medicine*, vol. 9, no. 8, pp. 894–898, 2008.

[17] Asela S. Karunajeewa, Udantha Ranjith Abeyratne, and Craig Hukins, "Silence-breathing-snore classification from snore-related sounds," *Physiological Measurement*, vol. 29, no. 2, pp. 227–243, 2008.

[18] Ali Azarbarzin and Zahra M. K. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals.," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1156–1162, May 2011.

[19] Lawrence J. Epstein, David Kristo, Patrick J. Strollo, Norman Friedman, Atul Malhotra, Susheel P Patil, Kannan Ramar, Robert Rogers, Richard J. Schwab, Edward M. Weaver, Michael D. Weinstein, and Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine, "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults.," *Journal of Clinical Sleep Medicine*, vol. 5, no. 3, pp. 263–276, June 2009.

[20] Conrad Iber, Sonia Ancoli-Israel, Andrew Chesson, and Stuart F. Quan, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, American Academy of Sleep Medicine, 2007.