Spectral and Bispectral Analysis of Awake Breathing Sounds for Obstructive Sleep Apnea Diagnosis

by

Davood Karimi

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

Abstract

The goal of this study was to investigate the potential of breathing sounds recorded during wakefulness for Obstructive Sleep Apnea (OSA) screening and severity estimation. Breathing sounds were recorded from 189 subjects in supine and sitting postures during nose and mouth breathing. Features were extracted from power spectrum and bispectrum of the signals. Data from 70 subjects were used for training. Validation accuracy, specificity, and sensitivity for non-OSA and OSA groups were 78%, 77%, and 82%, respectively. Screening based on six OSA risk factors was less accurate. Parallel classification by both breathing sound features and risk factors had high sensitivity (94%). OSA severity estimation, by classifying subjects into three classes of OSA severity, achieved a maximum validation accuracy of 71%. The results demonstrate the potential of breathing sounds for OSA screening. The proposed method can lead to significant improvements in efficient use of resources such as sleep laboratories.

Acknowledgments

I am very grateful to my supervisor, Prof. Zahra Moussavi, for her help, support, and guidance during the two years of my M.Sc. program.

I am very thankful to Dr. Danny Mann and Dr. Sherif Sherif for serving on my thesis committee and for their useful comments and suggestions.

This project was funded through an IPS scholarship by Natural Sciences and Engineering Research Council (NSERC) and SagaTech Electronics Inc. I am very grateful for their support.

I am also very grateful to the staff of TRLabs Winnipeg, especially to my supervisor Dr. Sergio Camorlinga, for their support and help.

The data recording for this thesis was performed at the Sleep Disorder Centre in Misericordia Health Centre in Winnipeg. I am very grateful for their help, support, and patience.

I am very grateful to my family, especially my parents, for their patience and for their constant prayers.

I thank God the Almighty for his mercy and forgiveness throughout my life.

Contents

Abstract i

Acknowledgments ii

Table of contents iii

List of Tables vi

List of Figures vii

1 Introduction 1

1.1 Obstructive Sleep Apnea 1

1.2 Pathophysiology of OSA 2

1.3 OSA diagnosis 4

1.4 OSA Risk Factors 6

1.4.1 Male sex 7

1.4.2 Obesity 7

1.4.3 Aging 8

1.4.4 Smoking 8

1.4.5 Other risk factors 8

1.5 Objectives 9

1.6 Organization of the Thesis 9

2 Methods 11

2.1 Data Collection 11

2.1.1 Subjects 11

2.1.2 Data recording 12

2.2 Data Analysis 14

2.2.1 Data Pre-processing 14

2.2.2 Feature Extraction 15

2.2.3 Feature Selection 25

2.3 Classification 33

2.3.1 Classification Based on Breathing Sound Features 33

2.3.2 Classification Based on Risk Factors 39

3 Results and Discussion 43

3.1 Classification Results for OSA Screening 43

3.1.1 Classification Based on Breathing Sound Features 43

3.1.2 Classification Based on Risk Factors 47

3.2 Classification Results for OSA Severity Estimation 49

3.2.1 Classification Based on Breathing Sounds 49

3.2.2 Classification Based on Risk Factors 50

4 Discussion 54

5 Conclusion and Recommendations for Future Work 65

5.1 Conclusion 65

5.2 Recommendations for future work 67

Bibliography 70

Appendix A Description of features in the best feature combinations for OSA screening 83

and OSA severity estimation

A.1 Best set of features for OSA screening 83

A.2 Best set of features for OSA severity estimation 84

List of Tables

2.1. Anthropometric information (mean ± standard deviation) for the subjects. 12

2.2. Breakdown of the number of subjects used in classification for OSA screening. 35

2.3. Breakdown of the number of subjects in classification for OSA severity estimation. 39

3.1. Summary of classification results for the test dataset including 70 subjects with AHI<10 and 17 subjects with AHI>20. 45

3.2. Summary of classification results for the subjects in the middle group ($10 < AHI \leq 20$). 46

3.3. The results of the three-class classification based on breathing sound features on the training dataset. 49

3.4. The results of the three-class classification based on breathing sound features on the testing dataset. 50

3.5. The results of the three-class classification based on breathing sound features for subjects with $5 \leq AHI < 10$ and $25 \leq AHI < 30$. 50

3.6. The results of the three-class classification based on estimated relative OSA risk on the training dataset. 52

3.7. The results of the three-class classification based on estimated relative OSA risk on the testing dataset. 53

3.8. The results of the three-class classification based on the estimated OSA risk for subjects with $5 \leq AHI < 10$ and $25 \leq AHI < 30$. 54

List of Figures

2.1. The experimental setup used for data recording. 13

2.2. The spectrogram of a typical tracheal breath sound signal recorded in this study (a), and the logarithm of the sound variance used to separate individual inspiration and expiration phases (b). 15

2.3. An example plot of the logarithm of absolute value of the estimated bispectrum. 20

2.4. Feature selection steps; "n" represents the number of features at each step. 25

2.5. (a) A sample plot of normalized feature values; the feature depicted in this plot is the signal power in the band of 450-600 Hz for inspiration phase of nose breathing in supine posture. The data for OSA and non-OSA subjects are shown on two different levels to avoid clutter. (b) The ROC curve for the feature plotted in part (a). The area under the ROC curve in this case is equal to 0.654. 27

2.6. Schematic demonstration of a minimum-distance classifier based on one feature. The data for OSA and non-OSA subjects are shown on two different levels to avoid clutter. The value of the feature for the subject to be classified is shown with the solid blue circle. 33

3.1. Relative OSA risk for the 70 subjects in the training dataset. The red horizontal line shows the threshold used for classification. 47

3.2. Relative OSA risk for the 87 subjects in the testing dataset. The red horizontal line 48

shows the threshold used for classification.

3.3. Relative OSA risk for the 55 subjects in the training dataset for estimation of OSA severity based on risk factors. The red horizontal lines show the thresholds used for classification. 51

3.4. Relative OSA risk for the 98 subjects in the testing dataset for estimation of OSA severity based on risk factors. The red horizontal line shows the threshold used for classification. 53

Chapter 1

Introduction

1.1 Obstructive Sleep Apnea

Sleep is an essential requirement for overall physical and mental health and well-being. Adequate proper sleep is necessary for daytime alertness and optimum productivity. Not only does it have a vital role in the proper functioning of the nervous system, it is also essential for the normal functioning of the body's immune system and ability to fight disease and sickness [1].

Obstructive sleep apnoea (OSA) is a common sleep disorder, characterized by repetitive pharyngeal collapses resulting in extended pauses in breathing (apnea) or instances of abnormally low breathing flow (hypopnea) during sleep. There is extensive evidence suggesting that untreated OSA may result in neurocognitive impairments and cardiovascular complications including hypertension, heart failure, stroke, excessive daytime sleepiness, as well as the increased risk of occupational and traffic accidents [2-5].

Severity of OSA is quantified by apnea hypopnea index (AHI), which is the number of apnea and hypopnea events per hour of sleep. An apnea is defined as complete cessation of respiratory

airflow for at least 10 s, whereas a hypopnea is characterized by a decrease of at least 50% in airflow that lasts at least 10 s [6]. Apnea/hypopnea events usually result in a significant drop (>4%) in blood's Oxygen saturation level (SaO2) [6]. In OSA, breathing is interrupted despite respiratory effort, because the airway is physically blocked. In another type of apnea, known as central sleep apnea (CSA), the effort to breathe is too weak or absent. There is also a third type of apnea, mixed sleep apnea, which is a combination of OSA and CSA. However, OSA is by far the most common type of sleep apnea [7].

OSA is highly prevalent; the general estimate is that between 2-7% of women and 4-14% of men above the age of 30 suffer from moderate or severe OSA [8-10]. Yet, studies have also shown that OSA is massively underdiagnosed. For example, a major study in the United States found that in a population with access to a sleep disorders clinic, more than 80% of middle-aged men and women who suffered from moderate or severe OSA had not been diagnosed [11]. In addition to its negative impacts on the patient's overall health and quality of life and increased risk of accidents, undiagnosed OSA is very costly to the medical system. A study has estimated that mean annual medical cost of an undiagnosed OSA patient is approximately twice that of a non-OSA individual, causing an estimated $3.4 billion in additional annual medical costs in the U.S. in 1999 [12].

1.2 Pathophysiology of OSA

Unlike most mammals, the human pharyngeal airway does not have rigid skeletal support. This is probably due to the development of speech in the evolution of human species [13]. As a result, the human upper airway is kept open by muscle activation and soft tissue. The patency of the upper airway is a function of complex balance of the forces that tend to close it and those that tend to keep it open. The former include the negative airway pressure during inspiration

and positive pressure outside the airway as a result of extra fat surrounding the upper airway [14]; the latter include the action of dilator muscles such as genioglossus muscle [15].

Imaging studies have confirmed that individuals with OSA, in general, have a smaller upper airway size compared to healthy controls at the same pharyngeal dilator muscle activity [16]. More than 20 muscles in the upper airway are responsible for maintaining its patency. Precise control and coordination of these muscles is necessary for speech, swallowing, breathing, and other tasks. Therefore, activation of these muscles is accurately controlled to meet the demands of specific tasks, but their activation is also influenced by sleep [17]. During wakefulness, when a negative pressure develops in the upper airway (e.g., due to inspiration), dilator muscles will respond within milliseconds, with a response that is linearly proportional to the negative pressure [18]. Because of the anatomical abnormalities of the upper airway mentioned above, the dilator muscles of individuals with OSA are significantly more active compared to healthy individuals [19]. This increased muscle activity has two reasons. First, because of the narrower upper airway, OSA individuals will require a larger rate of airflow that will, in turn, result in a larger negative airway pressure. Since the activity of the dilator muscles is proportional to the negative pressure, this will lead to increased muscle activity. Second, even in the absence of the negative pressure, the activity of these muscles in OSA individuals is higher than non-OSA individuals. Even though the mechanisms behind this are not completely understood, it is thought to be due to plasticity of the neural system [20].

The patency of the upper airway is easily maintained during wakefulness even in individuals with severe OSA. During sleep, on the other hand, the response of the dilator muscles to negative airway pressure significantly decreases. This decrease is seen in OSA individuals and healthy individuals alike [20]. Even though the details of the mechanisms involved are still debated, a

loss or decrease of neuromuscular reflexes is the likely reason [21]. Obviously, if an individual's upper airway patency during wakefulness depends on the reflex activation of dilator muscles, that individual will be at risk of airway collapse during sleep when the reflex-driven activation of those muscles are lost or weakened. Even though this is believed to be the main mechanism behind OSA [20], other factors and mechanisms may play important roles. For example, it is believed that reduced lung volume during sleep will lead to smaller longitudinal traction of the upper airway and, therefore, increased risk of closure [22]. Studies have also shown that the loop gain of the ventilatory control system in OSA patients is significantly larger than that of healthy controls. Increased loop gains mean a less stable ventilatory control system that may contribute to OSA [23].

1.3 OSA diagnosis

The current standard method for OSA diagnosis is overnight polysomnography (PSG) in a sleep laboratory. PSG includes simultaneous recording of electroencephalogram, electrooculogram, electromyogram of chin and anterior tibialis, respiratory airflow, electrocardiogram, pulse oximetry, and snoring sounds. Because of the high cost and time-consuming nature of PSG, many researchers have tried to develop simpler and faster diagnostic methods. This has led to increasing usage of simpler diagnostic modalities of unsupervised monitoring such as home sleep studies, which are now considered a part of commonly utilized tests in many sleep labs, with continuing need to further develop faster and easier-to-apply screening methods. Most of these methods essentially rely on a small subset of the signals used in PSG [24-26]. Some other methods have used short-time recordings of the nasal or oronasal airway pressure [27, 28], speech signals [29], and breath sounds [30-32]. The fact that a large number of people with OSA

are not diagnosed is partly because the current OSA diagnostic methods, such as PSG, are complicated and expensive.

Breath sounds signal analysis for OSA diagnosis has received considerable attention in recent years [19-21] because it is an inexpensive and non-invasive technology; breath sounds are easy to record, and are also representative of the OSA pathology. The rationale behind the use of breath sounds for OSA diagnosis lies in the anatomical and physiological differences of the upper airway between OSA and non-OSA individuals. Previous studies have demonstrated major anatomical abnormalities in people with OSA including a narrower oropharynx, thicker soft palate, and shorter intermaxillary space [22]. As mentioned before, this is compensated by the increased dilator muscle activity and more negative pharyngeal pressure during wakefulness [23]. However, these differences will alter the nature of airflow turbulence through the upper airway in OSA individuals compared to non-OSA individuals. A change in the airflow characteristics will, in turn, alter the breath sounds. If this hypothesis is true, these anatomical and physiological differences in the upper airway should be detectable by analysis of breath sounds during wakefulness.

The potential use of breath sounds analysis during sleep for OSA diagnosis has been noted as early as 1985 [24]. That study proposed a system for recording and analysis of tracheal respiratory sounds as a means to identify apneic events in infants. However, the suggested technique was only tested on rabbits. Later studies have successfully used breath sounds to detect the occurrence of apneic events and to classify them as central or obstructive apneic events during sleep [25, 26]. One of these studies used a pair of non-contacting microphones and achieved sensitivity of 99.2-100% and specificity of 52-82% in detecting apneic events in three separate clinical trials [25]. The other study used a microphone attached to a chestpiece

that was placed on the heart region of the chest [26]. Full-night respiratory sounds have also been used for identifying people with OSA. One such study used tracheal breath sounds and the blood oxygen saturation level to detect apnea and hypopnea events and estimate the AHI for 40 subjects [20]. The estimated AHI scores were very close to the AHI determined through full-night PSG, with a correlation coefficient of 0.96. Classifying the subjects into moderate or severe OSA and simple snorer groups resulted in sensitivity and specificity values of up to 100% and 96% respectively [20]. In another study using breath sounds of 60 subjects during sleep, the subjects were classified into two groups with a threshold of AHI=10 [27]; the sensitivity and specificity were reported as 96% and 91%, respectively.

On the other hand, there have been few efforts to diagnose OSA using breath sounds during wakefulness. One study compared the change in the intensity of tracheal breath sounds between supine and upright postures during wakefulness and found statistically significant differences between severe OSA patients (n=7) and healthy controls (n=8) [30]. However, in addition to the small number of subjects included in the study, no classification was performed. Another study used tracheal breath sounds recorded with different 1-min manoeuvres during wakefulness for OSA diagnosis. The results of this study with a relatively small number of subjects (n=52) showed a classification accuracy of 91% [32]. However, due to the small size of population, the classification and feature selection stages were not fully unbiased.

1.4 OSA Risk Factors

This section briefly describes major risk factors for OSA. These include male sex, obesity, aging, smoking, menopausal status, black race, and alcohol [20, 33]. Even though the mechanisms behind some of these risk factors, such as age and gender, are not entirely clear, studies have consistently shown that they are correlated with the risk of OSA [34, 35]. Physicians use these

risk factors, along with OSA symptoms such as snoring and witnessed apneas, to roughly estimate the OSA risk when deciding on whether a patient should be referred for comprehensive tests such as PSG.

1.4.1 Male sex

Men are at higher risk of OSA compared to women. This has been suggested to be, at least in part, due to hormonal influences [36]. Some researchers have suggested gender differences in the shape of the upper airway, genioglossal muscle activity, and pattern of fat deposition in the upper airway as possible causes [37]. However, there is no conclusive evidence to support any of these hypotheses [34]. The increased risk due to male sex reported by most studies is between 1.5 and 3 [38].

1.4.2 Obesity

Obesity is particularly important because it is the only major reversible risk factor and it is also becoming an epidemic, especially in developed countries. The effect of overweight and obesity on increased risk of OSA has been established by many studies [e.g. 39, 40]. It has been suggested that obesity affects breathing in several ways, including change in the shape of the upper airway, change in the function of the upper airway (e.g., by increasing its collapsibility), by reducing the functional residual capacity, and by increasing the oxygen demand of the body [41, 42]. Studies on the effect of obesity on OSA have relied on different body habitus measures including neck size [43], general obesity [44], and central obesity [45].

1.4.3 Aging

In general, OSA prevalence increases with age in midlife; however, this simple trend does not exist in childhood, adolescence, and older age [34]. Many studies have shown that OSA is highly prevalent in people older than age 65 years [e.g. 46, 47]. However, the OSA risk reaches a plateau after age 65 years [34].

1.4.4 Smoking

Smoking is often mentioned as a risk factor for OSA. However, there have been relatively few studies to investigate the effect of smoking. Most of these studies have found a significant positive correlation between smoking and OSA [48, 49]. There are also studies that have found opposite results. For example, a study on a large population (n=6,440) found a negative association between smoking and OSA [50]; in a model that adjusted for age and BMI, current smokers had significantly fewer respiratory disturbance events compared to never smokers. The exact mechanisms by which smoking affects OSA are not clear. Some researchers have suggested that smoking increases the risk of OSA through an increase in sleep instability and airway inflammation [34]. Other studies have suggested a rebound effect, in which the effects of nicotine that lead to the increased upper airway muscle tone are reversed during night, resulting in a decreased muscle activity and increased OSA risk [51].

1.4.5 Other risk factors

There are other OSA risk factors that are not considered in this study. These are minor risk factors that are briefly described here. Alcohol consumption is a manageable risk factor. Alcohol consumption increases nasal and pharyngeal resistance during wakefulness [52]. Therefore, it is reasonable to expect that it may increase the risk of OSA. In general, most studies have

confirmed this hypothesis [e.g. 53, 54]. However, there have also been some studies that did not find a significant association between alcohol consumption and OSA [55]. Ethnicity has also been shown to be a significant factor. For example, it has been found that OSA is more prevalent among African Americans compared to white people, especially at younger ages [56]. Studies have also found an increase in frequency of apnea in postmenopausal women compared with premenopausal women [57].

1.5 Objectives

The objectives of the study can be listed as follows:

I. To develop a classification algorithm for screening of OSA patients based on short recordings of breathing sounds during wakefulness.
II. To evaluate the power of clinical risk factors for OSA screening and compare a diagnostic classification based on risk factors with the proposed method of the objective 1 based on the breathing sounds.
III. To explore the potential of breathing sounds features for estimation of the severity of OSA.

1.6 Organization of the Thesis

The thesis is organized as follows:

Chapter 2 explains the methods followed to collect and process the breathing sound signals. It provides the details of the features and how they were extracted from breathing sounds. Feature reduction and feature selection methods are explained, followed by classifier development and testing. Classification based on OSA risk factors is also explained.

Chapter 3 presents the results of the classification based on breathing sounds and compares them with a classification based on risk factors.

Chapter 4 discusses the results and compares them with current OSA diagnostic methods. The possibilities of combining the classification based on breathing sound features with classification based on risk factors to achieve improved classification accuracy are also investigated. The results of OSA severity estimation based on breathing sound features and risk factors are also discussed.

Chapter 5 presents the conclusions of the thesis and suggestions for future research.

Chapter 2

Methods

2.1 Data Collection

2.1.1 Subjects

One hundred and eighty nine subjects, who had been referred to the full-night PSG, participated in this study. The study was approved by the Biomedical Research Ethics Board of the University of Manitoba, and all participants gave a written consent prior to collecting data. Data recording was performed at the Sleep Disorder Centre in Misericordia Health Centre in Winnipeg, MB. Our study was run in 10 minutes, while the subjects were awake before they proceeded for PSG assessment preparation. The AHI of the subjects was determined through the full-night PSG scoring by the skilled sleep lab technologists. For the purpose of developing an algorithm for OSA screening, the subjects were divided into three groups based on their AHI values: non-OSA group (subjects with $AHI < 10$), OSA group (subjects with $AHI > 20$), and middle group (subjects with $10 \leq AHI \leq 20$). Table 2.1 shows the description of each group and the number of subjects in each group as well as a summary of the anthropometric information for the subjects.

Table 2.1. Anthropometric information (mean ± standard deviation) for the subjects.

	age	BMI	weight (kg)	gender
Non-OSA $(AHI < 10);\ \ n = 105$	50.0 ± 14.8	31.3 ± 8.8	90.7 ± 27.3	48 F, 57 M
Middle group $(10 \leq AHI \leq 20);\ \ n = 32$	54.4 ± 12.9	32.3 ± 6.9	91.8 ± 21.2	13 F, 19 M
OSA $(AHI > 20);\ \ n = 52$	53.4 ± 11.9	35.2 ± 6.5	106.3 ± 22.5	7 F, 45 M

2.1.2 Data recording

Breathing sounds were recorded with an ECM77B Sony microphone. The microphone was inserted into a plastic chamber allowing 2 mm cone shape space between the microphone and the skin; it was placed on the suprasternal notch of the trachea using a double-adhesive disk. The subjects were asked to breathe at their maximum respiratory flow rate at two different postures: while sitting upright on a chair or on the edge of the bed, and while lying on their back on the bed. At each posture, the subjects were asked to first breathe through their nose with their mouth closed. Then, they were asked to wear a nose clip and breathe only through their mouth. Therefore, from each subject four breathing sound signals were recorded, each containing 4-5 full breath cycles (inspiration and expiration). People usually tend to extend the duration of their breath when they breathe deeply. To minimize the effect of this habit, the subjects were asked to coordinate their inspiration-expiration with up-down movement of the hand of the experimenter. The experimenter stood next to the bed and moved his hand up and down at a relatively constant timing. Using this strategy, all subjects breathed at almost equal flow rate of approximately 17 breaths per minute. The signals from the microphone were band-pass filtered to remove the frequency components below 0.05 Hz and above 5000 Hz, amplified,

and recorded on a laptop computer. The sampling frequency was set at 10240 Hz. Figure 2.1 shows the setup used for data recording.

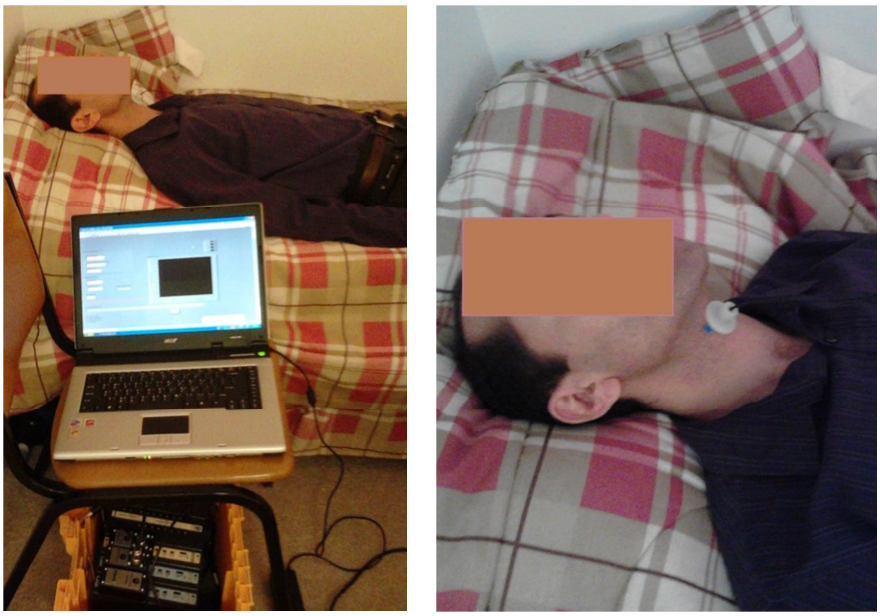


Figure 2.1. Tracheal breath sound recording.

After recording the breathing sounds, the subject's neck circumference was measured with a plastic tape measure to the nearest 1.0 cm and their Mallampati score was determined. Mallampati score, also known as Mallampati classification, is a score that is based on the visibility of uvula, faucial pillars, and soft palate and is used in anaesthesiology to predict the ease of intubation [58]. To determine the Mallampati score, the subject is instructed to open his/her mouth, while protruding their tongue. Mallampati score can range from 1 to 4 as follows [59]:

Class I: soft palate, tonsils, and uvula are completely visible,

Class II: soft palate and the upper portion of uvula are visible,

Class III: soft palate and the base of uvula are visible,

Class IV: soft palate is not visible

The subject's body mass index (BMI) was determined by measuring their height and weight at the same night that they completed the PSG test. The subject's age, gender, and smoking history were collected from a questionnaire that the subject completed prior to PSG.

2.2 Data Analysis

2.2.1 Data Pre-processing

Features were computed for each inspiration and expiration phase. As mentioned in the previous section, each signal contained 4-5 breath cycles. To ensure correct identification of inspiratory/expiratory phases, the first inspiratory phase in each signal was marked verbally during recording. The method introduced in [31] was used to detect the breath onsets and separate inspiration and expiration phases in each signal. In brief, the logarithm of the sound's variance (log-var) was computed in windows of 50 ms with 95% overlap (Figure 2.2). The valleys of the log-var signal are basically the breath onsets; all the detected valleys were checked manually to ensure accuracy. Knowing the first phase as inspiration and given the fact that the breathing was regular (alternating phases), we used the breath onsets and labeled all the breath phases as inspiration or expiration; these two respiratory phases were analyzed separately. After separating inspiration and expiration phases, the signal from each phase was divided by its standard deviation in order to minimize the effect of the difference in respiratory flow between the subjects and between the breathing cycles of each subject.

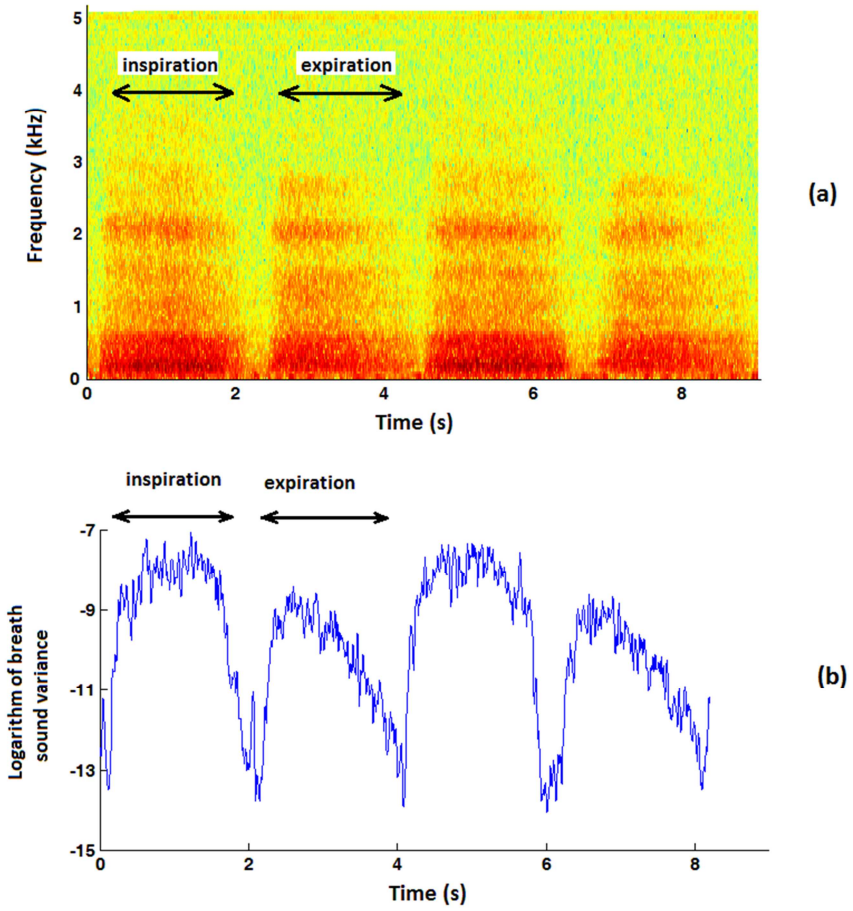


Figure 2.2. The spectrogram of a typical tracheal breath sound signal recorded in this study (a), and the logarithm of the sound variance used to separate individual inspiration and expiration phases (b).

2.2.2 Feature Extraction

Features were computed from the power spectrum and bispectrum of the signals. Power spectrum of stochastic signals has been used in various fields of engineering and science in the past 30 years, and is currently an essential signal processing tool in communication, speech, radar, biomedical, and other signal processing applications. However, power spectrum suppresses phase relationships between frequency components. The information in the power spectrum of a signal is identical to that contained in the autocorrelation function of the signal.

This information is sufficient for complete description of the signal only if the signal is Gaussian. Higher-order spectra are also used to detect and characterize nonlinear properties of signals and systems [60, 61]. Tests of Gaussianity and linearity introduced in [62] were run on all inspiration and expiration segments of the data collected in this study. An overwhelming majority of the signals did not pass these tests. Therefore, it was decided to extract characteristic features from the power spectral density (PSD) as well as the bispectrum of the signals.

The breathing sound signal from an inspiration or expiration phase in each of the 4 recordings was first normalized by its standard deviation as mentioned above. Tracheal breathing sounds are broad-spectrum signals covering a frequency range that starts from below $100\,Hz$ and extends up to and above $1500\,Hz$, usually with a significant drop in signal power approximately at $800\,Hz$ [63]. However, tracheal breathing sounds below $100\,Hz$ are often contaminated by various types of noise, such as heart sound and background noise. Furthermore, since the breathing sounds in this study were collected at maximum airflow, they had significant power up to $2500\,Hz$. Therefore, we estimated power spectrum density of each normalized signal in the frequency range of 100 Hz to 2500 Hz. The Welch estimation method with a Hamming window of 80 ms length and 50% overlap was used. The choice of the window length was made by trial and error. Windows of length 80 ms provided a good compromise between the smoothness of the estimated PSD and its frequency resolution.

From the estimated power spectrum, the following features were computed:

i. Signal power- it gives the power of the signal in the frequency band in consideration [64]:

$$\sum_{f=f_u}^{f=f_l} P(f)\Delta f, \tag{2.1}$$

where $P(f)$ is the estimated PSD and f_u and f_l represent the upper and lower limits of the given frequency band.

ii. Spectral centroid – it finds the weighted average frequency of the area under the PSD for a given frequency band [65]. Therefore, this feature can identify the location of major peaks, if any.

$$\frac{\sum_{f=f_u}^{f=f_l} f\,P(f)\Delta f}{\sum_{f=f_u}^{f=f_l} P(f)\Delta f} \tag{2.2}$$

iii. Spectral bandwidth – it finds the weighted average of the squared distance between different frequency components and the spectral centroid, SC, with the weight being the value of the estimated PSD at each frequency [66].

$$\frac{\sum_{f=f_u}^{f=f_l} (f-SC)^2.P(f)\Delta f}{\sum_{f=f_u}^{f=f_l} P(f)\Delta f} \tag{2.3}$$

iv. Spectral flatness – also called tonality coefficient; it quantifies how tone-like a signal is, as opposed to being noise-like [67]. For a completely flat power spectrum, i.e. white noise, it evaluates to 1. As shown by the following equation, spectral flatness is computed as the ratio of the geometric mean to the arithmetic mean of the PSD.

$$\frac{\left(\prod_{f=f_u}^{f=f_l} P(f)\right)^{\frac{1}{fu-fl}}}{\frac{1}{fu-fl}\sum_{f=f_u}^{f=f_l} P(f)\Delta f} \tag{2.4}$$

v. Crest factor –it is another measure of tonality of the signal. In other words, it can be used to distinguish between wideband signals (with smaller crest factor) and narrowband signals (with a larger crest factor) [66].

$$\frac{\max(P(f))}{\frac{1}{f_u - f_l}\sum_{f=f_u}^{f=f_l} P(f)\Delta f} \tag{2.5}$$

All these features were computed for the entire frequency band used in power spectrum estimation, i.e. 100-2500 Hz, as well as for six sub-bands: 100-150 Hz, 150-450 Hz, 450-600 Hz, 600-1200 Hz, 1200-1800 Hz, and 1800-2500 Hz. The choice of these sub-bands was mainly through visual inspection of the shape of the estimated PSD and its apparent differences between non-OSA and OSA subjects. Furthermore, the relative power in each of the sub-bands was computed by dividing the signal power in that sub-band by the power in the band of 100-2500 Hz.

The bispectrum of a stochastic signal or random process is defined as the Fourier transform of its third-order cumulant as follows [60]:

$$C_3(f_1, f_2) = \sum_{\tau 1=-\infty}^{+\infty} \sum_{\tau 2=-\infty}^{+\infty} c_3(\tau_1, \tau_2) \exp\{-2\pi j(f_1\tau_1 + f_2\tau_2)\}, \tag{2.6}$$

where $c_3(\tau_1,\tau_2)$ is the third-order cumulant. For a zero-mean signal, the third-order cumulant is equal to the third-order moment defined by:

$$c_3(\tau_1, \tau_2) = m_3(\tau_1, \tau_2) = E\{X(K)X(K+\tau_1)X(K+\tau_2)\} \tag{2.7}$$

We used the conventional direct method of bispectrum estimation, which is an approximation of bispectrum for the time series with limited available samples [60]. This method can be implemented using the following steps:

1. The signal $x[n], n = 0,1, \dots, N-1$ is divided into K overlapping segments of length M. We used windows of length $80\ ms$ with 50% overlap. Each of these segments is denoted by $x^{(k)}[n], k = 0,1, \dots, K-1$.
2. Computed the zero-mean signal segments by subtracting the mean of the signal segment.
3. Compute the discrete Fourier transform (DFT) of the segment multiplied by a Hamming window:

$$X^{(k)}(\lambda) = \frac{1}{M}\sum_{m=0}^{M-1} x[n]w[n]\exp(-i2\pi n\lambda/M) \tag{2.8}$$

4. Compute the raw estimate of bispectrum as:

$$\widehat{C_3}^{(k)}(\lambda_1, \lambda_2) = X^{(k)}(\lambda_1)X^{(k)}(\lambda_2)X^{(k)}(\lambda_1 + \lambda_2) \tag{2.9}$$

5. Estimate of the bispectrum will be obtained by averaging the estimates from each segment:

$$\widehat{C_3}(\lambda_1, \lambda_2) = \frac{1}{K}\sum_{k=0}^{K-1} \widehat{C_3}^{(k)}(\lambda_1, \lambda_2) \tag{2.10}$$

Since the bispectrum is symmetric across multiple axes due to the symmetric nature of the cumulant, we only considered the non-redundant region which is defined as follows [61]:

$$\Omega = \{(f_1, f_2) |\ \ 0 \le f_1 \le f_n,\ \ 0 \le f_2 \le f_1,\ \ 2f_1 + f_2 \le 2f_n\}, \tag{2.11}$$

where $f_n$represents the Nyquist frequency. Furthermore, we limited our analysis of the bispectrum to the frequencies ranging from 100-2600 Hz in every dimension of the bispectrum with a frequency resolution of 5 Hz. Figure 2.3 shows an example of estimated bispectrum.

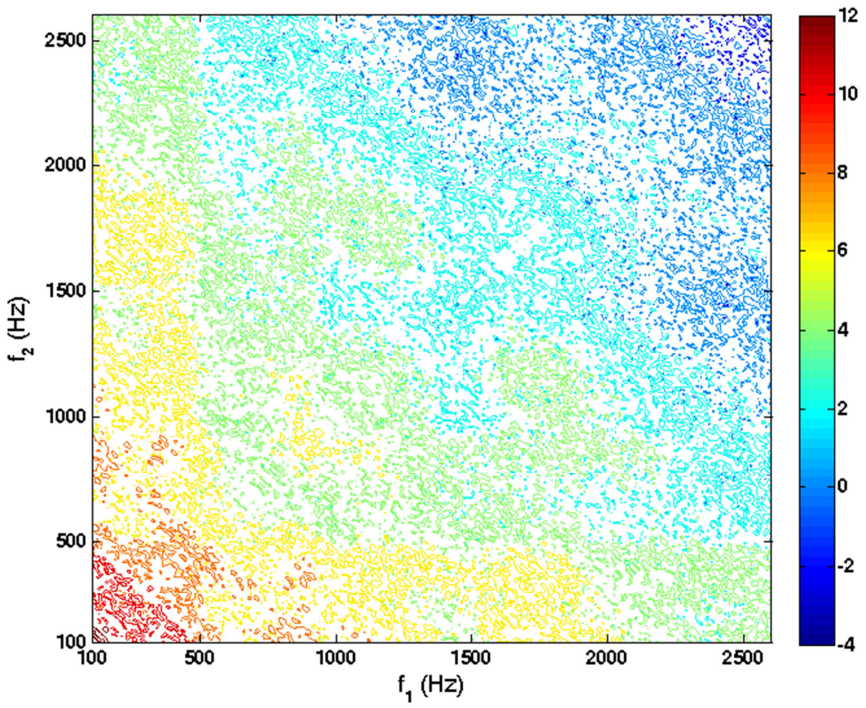


Figure 2.3. An example plot of the logarithm of absolute value of the estimated bispectrum.

From the estimated bispectrum, the following features were extracted. These features have been successfully applied to biomedical signals in previously published research [68-71].

i. Bispectral invariant parameter, $P(a)$ proposed in [72], where the authors also prove that these features are invariant to translation, amplification, and DC level.

$$P(a) = \text{atan}\left(\frac{I_i(a)}{I_r(a)}\right) \tag{2.12}$$

where

$$I(a) = I_r(a) + i\, I_i(a) = \int_{f_1=0^+}^{\frac{1}{1+a}} C_3(f_1, af_1)\, df_1 \tag{2.13}$$

The non-redundant region corresponds to a range of $0 < a \leq 1$. We estimated $P(a)$ on radial lines with a slope of 1° to 45° with 1° intervals. The integration was approximated

using the technique suggested in [68]. Specifically, the integral in (2.13) is approximated as

$$I(a) = \sum_{k=1}^{\left\lfloor \left(\frac{N}{2}-1\right)/(1+a) \right\rfloor} C_3(k, ak). \tag{2.14}$$

The bispectrum is interpolated as:

$$C_3(k, ak) = p\, C_3(k, \lfloor ak \rfloor) + (1-p) C_3(k, \lceil ak \rceil), \tag{2.15}$$

where $p = ak - \lfloor ak \rfloor$, $\lfloor x \rfloor$ represents the largest integer contained in x, and $\lceil x \rceil$ represents the smallest integer that is larger than x.

ii. Average magnitude of the bispectrum in the non-redundant region, computed as:

$$M_{avg} = \frac{1}{N} \sum_{\Omega} |C_3(f_1, f_2)|, \tag{2.16}$$

where Ω indicates the non-redundant region [70].

iii. Average power of the bispectrum in the non-redundant region:

$$P_{avg} = \frac{1}{N} \sum_{\Omega} |C_3(f_1, f_2)|^2 \tag{2.17}$$

iv. Normalized bispectral entropy:

$$E_1 = -\sum_n p_n \log p_n, \tag{2.18}$$

where

$$p_n = \frac{|C_3(f_1,f_2)|}{\sum_{\Omega} |C_3(f_1,f_2)|}. \tag{2.19}$$

v. Normalized bispectral squared entropy:

$$E_2 = -\sum_n q_n \log q_n, \tag{2.20}$$

where

$$q_n = \frac{|C_3(f_1,f_2)|^2}{\sum_\Omega |C_3(f_1,f_2)|^2}. \tag{2.21}$$

vi. Sum of logarithmic amplitudes:

$$H_1 = \sum_\Omega \log(|C_3(f_1,f_2)|) \tag{2.22}$$

vii. Sum of logarithmic amplitudes of diagonal elements of the bispectrum:

$$H_2 = \sum_\Omega \log(|C_3(f,f)|) \tag{2.23}$$

viii. First-order moment of the logarithmic amplitudes of diagonal elements:

$$H_3 = \sum f.\log(|C_3(f,f)|) \tag{2.24}$$

ix. Second-order moment of the logarithmic amplitudes of diagonal elements:

$$H_4 = \sum (f - H_3)^2.\log(\,|C_3(f,f)|) \tag{2.25}$$

x. Phase entropy of the estimated bispectrum:

$$P_e = \sum_n p(\psi_n) \log\big(p(\psi_n)\big), \tag{2.26}$$

where

$$p(\psi_n) = \frac{1}{N}\sum_\Omega 1\big(\phi(C_3(f,f) \in \psi_n)\big), \tag{2.27}$$

$$\psi_n = \{\phi | -\pi + 2\pi n/M \le \phi \le -\pi + 2\pi(n+1)/M\}.$$

xi. The median bifrequency - as the name suggests, median bifrequency is the frequency at which the area under the bispectrum is equal on both sides. Median bifrequencies were calculated along each frequency dimension of the non-redundant region of the bispectrum using the following approach:

a. The sum of all values of the bispectrum for all bifrequencies in the non-redundant region was calculated.

b. The value of f_1 was set to the smallest value in the non-redundant region.

c. The sum of bispectral values for all bifrequencies (f_1, f_2) in the non-redundant region was calculated.

d. If the sum calculated above was greater than half of the sum calculated in step (a), then the value of f_1 was taken as the median frequency for the first dimension. Otherwise, f_1 was incremented by $5Hz$ (the frequency resolution of the estimated bispectrum) and steps (c) and (d) were repeated. The sum in step (c) was calculated accumulatively with previous runs.

Similar steps were repeated for the second dimension of the bispectrum to obtain the median bifrequency in the f_2 direction.

xii. Average amplitude of the bispectrum over equal and non-overlapping regions in the non-redundant region was computed. Specifically, the frequency band of 100-2600 Hz in each of the $f_1$and f_2 frequency axes was divided into 10 equal non-overlapping sub-bands, each extending 250 Hz in frequency. This partitioned the non-redundant region into a set of square-shaped regions (triangle-shaped for regions on the diagonal). The average amplitude of the bispectrum over each of these regions was computed.

For each of the four breath sound signals, after computing the above features for individual inspiration and expiration phases, they were averaged to give one feature value for inspiration and one for expiration. Studies have shown that the power spectral features of tracheal breathing sounds are highly variable between subjects but they show very little variability within the same subject [73]. Therefore, by averaging the feature values for the same subject, we hope to obtain a more representative and less variable feature value. Furthermore, the difference in feature values between nose and mouth breathing as well as the difference between upright and supine breathing were computed.

As mentioned in the previous chapter, the idea of using breathing sounds for OSA screening is based on the assumption that the anatomical differences in the upper airway between OSA and non-OSA subjects lead to differences in the turbulent air flow through the upper airway. It is reasonable to expect that these differences would change between supine and upright postures and between nose breathing and mouth breathing. This is the rationale behind computing the difference of features between supine and upright postures and between nose breathing and mouth breathing. In other words, the difference of a feature between, say, nose breathing and mouth breathing may bring forth patterns that are not present in the feature from nose breathing or mouth breathing alone.

In addition, all features were normalized such that each had a mean of zero and a unity standard deviation. Normalization of the features is necessary because, as described in the next section, feature selection and classification steps include computation of the difference between feature values and comparing them across the features. Therefore, normalization of features ensures that all features are given equal weight in feature selection and classification steps.

2.2.3 Feature Selection

A total of 912 features were computed for each subject. Because this was a very large number and the features were highly correlated, it was necessary to identify a small subset of features with the highest classification power. This was done in several steps as described sequentially in this section. These steps are shown schematically in Figure 2.4.

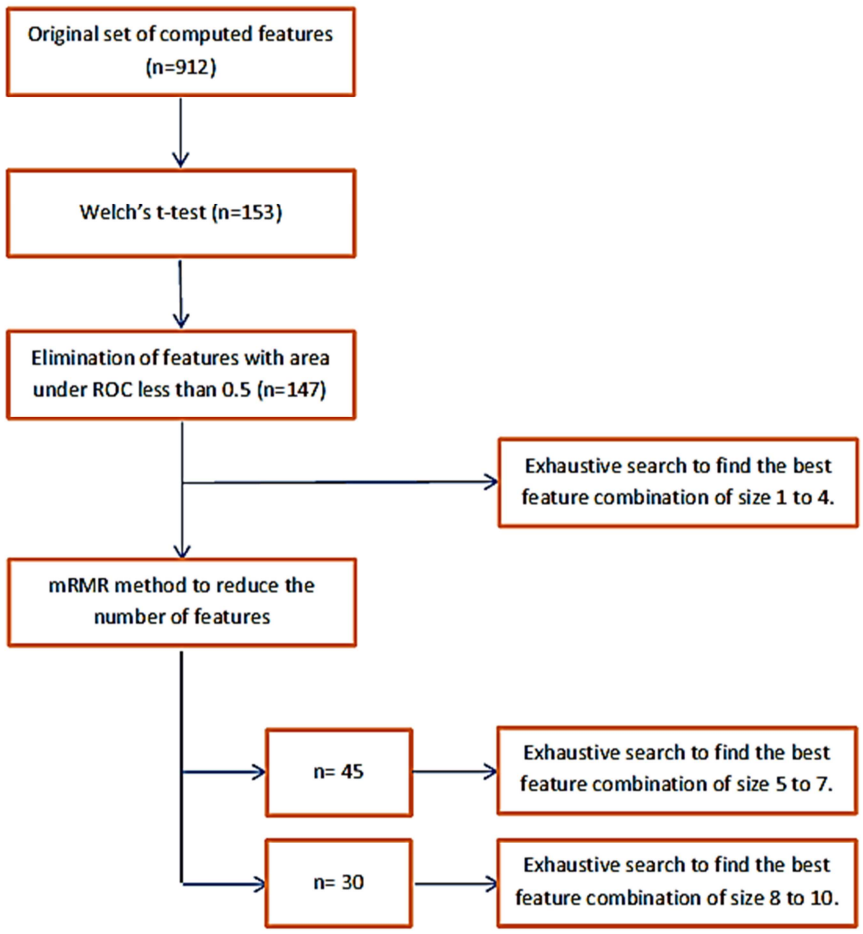


Figure 2.4. Feature selection steps; "n" represents the number of features at each step.

In statistical classification, it is usually recommended to use between 67% and 75% of the entire data for training and the remainder of the data for performance evaluation [74, 75]. Deciding on the amount of data to use for training and testing steps was more complicated in this study

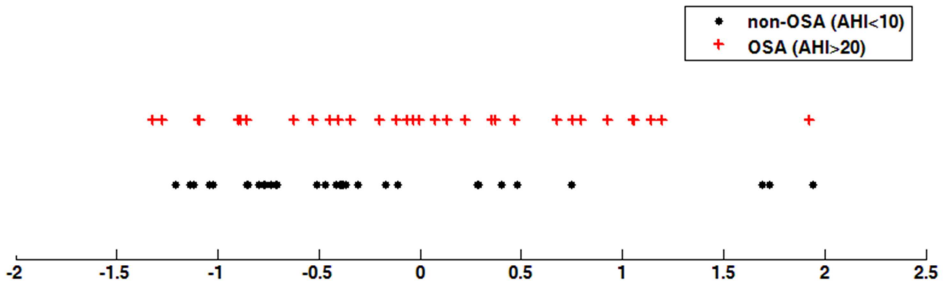
because the number of subjects in different groups were highly unequal (Table 2.1). We decided to use the data from 35 non-OSA (AHI<10) and 35 OSA subjects (AHI>20) for feature selection steps. By using 35 subjects from non-OSA and severe-OSA groups, we hoped to be able to select the best feature subsets in terms of classification power, while leaving enough data in each group for testing step.

A. **Welch's t-test.** First, Welch's t-test was used to identify those features that were statistically different between the two groups of subjects. For this step, data from 20 of the non-OSA (AHI<10) and 15 of the OSA subjects (AHI>20) were used. A total of 153 features were significantly different at p= 0.05 level.

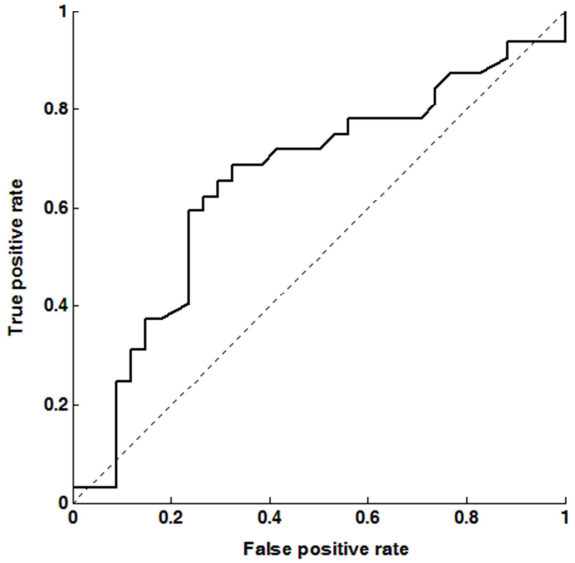
B. **The receiver operating characteristic (ROC) curve.** Statistical tests, such as the t-test used in step A, provide statistical evidence regarding the difference in the mean value of a feature between two or more classes. Even though this is an effective method for discarding less important features, the features that pass these tests do not necessarily possess a discriminative power. In other words, although the mean values may differ significantly between the classes, the spread of the feature values around the mean values can be very large. Therefore, it is essential to examine the discriminative power of the features that pass the statistical test. One of the simplest and most widely used criteria for this purpose is the area under the receiver operating characteristics (ROC) curve [74]. In general, the ROC curve is the plot of the true positive rate versus the false positive rate of a binary classifier as its threshold is varied.

   For each of the 153 features that had passed the t-test, a simple threshold classifier was built to classify the 35 subjects with AHI<10 and 35 subjects with AHI>20. Then, the area under the receiver operating characteristics (ROC) curve was computed by changing the

value of the threshold. Figure 2.5 shows a plot of the feature values for a particular feature along with the ROC curve for this feature. A total of 6 features that had an area under the ROC curve less than 0.50 were removed from the list, leaving 147 features for the next steps.



(a)



(b)

Figure 2.5. (a) A sample plot of normalized feature values; the feature depicted in this plot is the signal power in the band of 450-600 Hz for inspiration phase of nose breathing in supine posture. The data for OSA and non-OSA subjects are shown on two different levels to avoid clutter. (b) The ROC curve for the feature plotted in part (a). The area under the ROC curve in this case is equal to 0.654.

C. Minimal-redundancy-maximal-relevance (mRMR) criterion. Usually an exhaustive search is best to select the most characteristic features for classification. However, an exhaustive search of all possible feature combinations, from among the 147 features of step B, is very time consuming and computationally expensive; in this case, it was only possible for sets of size up to 4. In order to choose features sets of size 5 or larger through exhaustive search, it was necessary to significantly reduce the number of features from 147. For this purpose, the mRMR method introduced in [76] was employed. This method is a filter-type feature selection method [74] that selects features with maximum relevancy to the class labels but also with minimum redundancy among themselves.

The mRMR method is based upon the Max-Dependency criterion, but makes simplifying assumptions so that the method can be applied to practical problems with limited data. In general, the goal of feature selection is to select a subset S of m features from among the n features ($m < n$) such that these m features have the largest joint dependency on the target class c. This is known as Max-Dependency criterion and is written as:

$$max\ \ D(S, c) \tag{2.28}$$

where $D(S, c)$, the dependency, is defined as:

$$D(S, c) = D(\{x_i, i = 1, \dots, m\}, c) = I(\{x_i, i = 1, \dots, m\}; c). \tag{2.29}$$

In the above equation, $I(x, y)$ is the mutual information between the two variables x and y and is defined in terms of their marginal and joint probability distributions as follows:

$$I(x; y) = \iint p(x, y) \log \frac{p(x,y)}{p(x)p(y)}\ dx\ dy \tag{2.30}$$

A feature selection approach based on Max-Dependency criterion starts by finding the best single feature (the feature with largest dependency with the class label) and incrementally adds more features until the desired number of features are found or some other condition is satisfied. Having found a feature set of size $m-1$, a feature set of size m is found by adding the feature that results in largest increase in the dependency $D(S_m, c)$. This can be written as:

$$D(S_m, c) = \iint p(S_m, c) \log \frac{p(S_m, c)}{p(S_m)p(c)}\, dS_m\, dc$$

$$= \iint p(S_{m-1}, x_m, c) \log \frac{p(S_{m-1}, x_m, c)}{p(S_{m-1}, x_m)p(c)}\, dS_{m-1} dx_m\, dc \qquad (2.31)$$

$$= \int \dots \int p(x_1, \dots, x_m, c) \log \frac{p(x_1, \dots, x_m, c)}{p(x_1, \dots, x_m)p(c)}\, dx_1 \dots dx_m\, dc$$

Even though Max-Dependency is a powerful concept, its implementation requires estimation of the joint probability densities in the above equation. Accurate estimation of these probability density functions requires very large data sets. The amount of data needed for accurate estimation of these probability density functions grows exponentially with the desired number of features, m. Therefore, the Max-Dependency criterion as formulated above cannot be accurately implemented for most practical applications. An alternative to Max-Dependency is maximal relevance criterion (also known as Max-Relevance) that estimates $D(S_m, c)$ in (2.28) by computing the average of the mutual information between the individual features and the class label. In other words, Max-Relevance aims at maximizing $D(S, c)$ defined as:

$$D(S, c) = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c), \qquad (2.32)$$

where $|S|$ denotes the cardinality of set S. This definition overcomes the main difficulty with the Max-Dependency criterion because (2.29) requires estimation of the joint probability density function of two variables only, which can be estimated accurately with a modest amount of data. However, because Max-Relevance relies only on the mutual information of individual variables and the class label, it is likely that the features selected based on this criterion will have significant redundancy among themselves. To overcome this problem, the authors of [76] define the redundancy among the features as follows:

$$R(S) = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \tag{2.33}$$

In other words, redundancy of the set of features S is simply the average of the mutual information between the pairs of features in that set. The minimal-redundancy-maximal-relevance (mRMR) criterion is then defined by combining (2.29) and (2.30) in the following simple way:

$$\max \Phi(D, R), \quad \text{where} \quad \Phi(D, R) = D - R \tag{2.34}$$

In the incremental form, having selected a set of $m-1$ features, S_{m-1}, set of size m is found by adding the feature that maximizes the following function:

$$\max_{x_j \in X - S_{m-1}} \left[I(x_j; c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_i; x_j) \right] \tag{2.35}$$

Estimation of the mutual information terms $I(x_i; c)$ and $I(x_i; x_j)$ requires numerical estimation of marginal and joint probability density functions $p(c)$, $p(x_i)$, $p(x_i; c)$, and $p(x_i; x_j)$. In this study, features were continuous, whereas the class labels were discrete. Therefore, for estimation of the mutual information between features and

class labels, $I(x_i; c)$, the features must be discretized too and the probability density functions should be estimated in discrete form. For estimation of the mutual information between pairs of features, $I(x_i; x_j)$, on the other hand, all probability density functions can be estimated in continuous form, as explained below.

For estimating $I(x_i; c)$, $p(c)$ can be easily estimated by counting the number of subjects in each class in the training data. Estimation of $p(x_i)$and $p(x_i; c)$ were carried out by discretizing the feature values such that the double integral in (2.27) could be written as a double summation. The number of bins used to discretize the features is a critical parameter and needs to be determined by trial and error for each specific problem. The optimum number of bins depends on the amount of data available; with a larger dataset the number of bins can be increased to achieve a more accurate estimation. In this study, the same 70 subjects that had been used in step B were used for this step. Trial and error showed that approximately 12 bins were optimal. After discretizing the feature values and estimating $p(x_i)$ and $p(x_i; c)$, the mutual information between features and class labels, $I(x_j; c)$, was estimated using a discrete version of (2.30):

$$I(x_i; c) = \sum p(x_i; c) \log \frac{p(x_i;c)}{p(x_i)p(c)}, \tag{2.36}$$

where summation is carried over non-zero bins.

For estimating $I(x_i; x_j)$, $p(x_i)$ and $p(x_j)$ were estimated using the built-in Matlab functions to fit a nonparametric kernel-smoothing distribution. In order to estimate $p(x_i; x_j)$, Parzen-window density estimation with a 2D Gaussian kernel was used:

$$p(x_i; x_j) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{2\pi h^2} \exp\left\{ -\frac{\left(x_i - x_i^{(k)}\right)^2 + \left(x_j - x_j^{(k)}\right)^2}{2\,h^2} \right\}, \tag{2.37}$$

where $\{(x_i^{(k)}, x_j^{(k)}),\ k = 1, \dots, N\}$ is the set of the training data. The important parameter in this estimation is h, which should be chosen based on the range of feature values as well as the amount of data (i.e. N). An optimal value for h can be found by plotting $p(x_i; x_j)$ for different values of h. In this study, $h = 0.3$ was used. Finally, $I(x_i; x_j)$ is computed by numerical integration of (2.30).

D. Exhaustive search. The final step in feature selection was an exhaustive search of all combinations of a large set of features in order to find the set of features that resulted in the highest classification accuracy on the training data set. For example, to choose a subset of size 3, all possible combinations of 3 features from among the 147 features were considered. For each set, a classifier was built and its classification accuracy on the training data set was evaluated; then, the set of 3 features that resulted in the highest classification accuracy on the training data was selected. However, exhaustive search within the 147 features from step B was only computationally possible for sets of features of size less than 5. To select the best sets of features of size 5 or larger, the set of 147 features had to be reduced further so that an exhaustive search would be possible. This was performed using the mRMR method explained in step C. As mentioned in step C, mRMR sorts the features based on minimal-redundancy-maximal-relevance criterion. The top 45 features from this method were used to choose best feature subsets of size 5, 6, and 7 exhaustively; the top 30 features were used to choose best feature subsets of size 8, 9, and 10 exhaustively. Numbers 45 and 30 were chosen so that the exhaustive search to determine the best feature subsets of size 5 to 10 could be completed in less than 48 hours on a PC with a 1.73-GHz processor. As mentioned before, the “best” feature subsets were defined to be the subsets that resulted in the highest classification accuracy on the training data. The selected sets of features were

then used to evaluate the classification accuracy on the testing dataset. The same classification method was used in both training and testing steps, as explained in the next step.

2.3 Classification

2.3.1 Classification Based on Breathing Sound Features

The proposed classification method is based on building a separate minimum-distance classifier for each feature. Figure 2.6 shows such a classifier. In this example, the distance from the feature value of the subject to be classified to the mean of the OSA group is smaller than that of the non-OSA group; therefore, the subject is classified as OSA. After a subject is classified based on each feature separately, the final class label of that subject is assigned by the majority vote. For example, assuming five features are included in the classification, if a specific subject is classified as "non-OSA" by three of the features and as "OSA" by the remaining two features, the subject is then classified as "non-OSA".

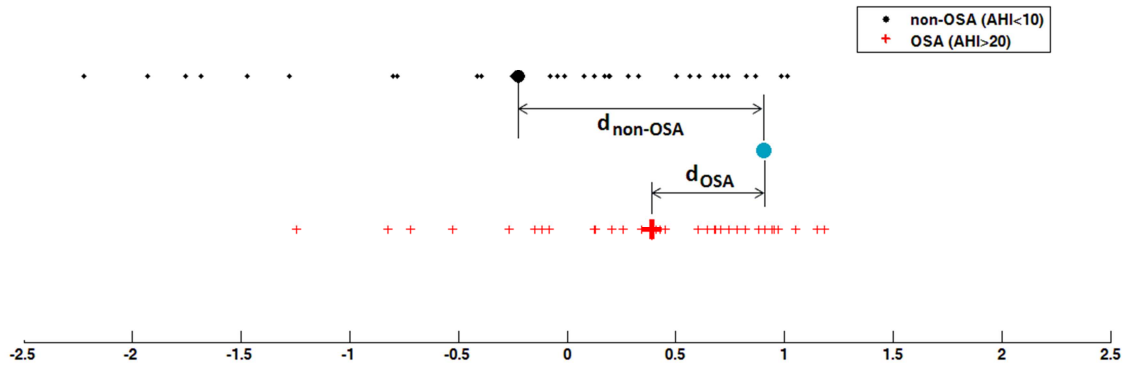


Figure 2.6. Schematic demonstration of a minimum-distance classifier based on one feature. The data for OSA and non-OSA subjects are shown on two different levels to avoid clutter. The value of the feature for the subject to be classified is shown with the solid blue circle.

It is possible that the number of votes is equal (due to having even number of features to votes and/or missing data for a particular feature). This may happen, for example, when four features are involved and a subject is classified as “non-OSA” by two features and as “OSA” by the other two. In such an event, the strengths of the decisions of the individual classifiers are considered. The strength of the decision of a classifier was defined as the difference in the distances of the feature value of the subject being classified with the mean feature value of the two classes. For example, the strength of the classification shown in Figure 2.6 is $d_{non-OSA} - d_{OSA}$. A feature, for which this difference is large, will provide a strong vote; whereas, a feature for which this difference is small, will provide a weak vote. The weighted sum of these votes will then be used to make the classification.

There were two main reasons for adopting this specific classification approach:

1. Our preliminary analysis showed that an optimal number of features for this study was more than 3. However, most classification methods require a very large training data in a high-dimensional feature space. For example, all classification methods that are based on finding a separating hyperplane suffer from this problem. Building a separate classifier for each feature overcomes this problem, because adding more features will not increase the dimensionality of the classifier and the involved classifiers are always in one dimension.
2. Many of the subjects in the training and testing data sets in this study did not have all the features due to noise that caused to exclude a signal partially or entirely from the analysis. The proposed classification approach allows classification of a subject even when some of the features involved in classification are missing. For example, if the

classification is based on 5 features and a specific subject is missing 2 of the features, that subject can still be classified based on the 3 existing features' votes.

The main objective of this study was to investigate the potential of breathing sound features for screening of OSA patients. For this purpose, the subjects were divided into three groups, as previously presented in Table 2.1.

Table 2.2 presents details of the number of subjects used in training and testing steps for this analysis. As shown in this table, 35 subjects with $AHI < 10$ and 35 subjects with $AHI \geq 20$ were used for training steps, leading to the selection of best sets of features of different sizes. Once the features were selected as explained in previous section, they were used to classify the remaining subjects with $AHI < 10$ ($n = 70$) and $AHI \geq 20$ ($n = 17$).

Table 2.2. Breakdown of the number of subjects used in classification for OSA screening.

Range of AHI	Total no. of subjects	No. of subjects in training step	No. of subjects in the first testing step	No. of subjects in the second testing step
$AHI < 10$	105	35	70	-
$10 \leq AHI < 15$	16	-	-	16
$15 \leq AHI < 20$	16	-		16
$AHI \geq 20$	52	35	17	-

Note that the test dataset was completely separate from the data used for feature selection and training the classifiers. The subjects with $10 \leq AHI < 20$ were also divided into two groups with a threshold of A$HI = 15$ and were classified. In all classification steps, a leave-one-out strategy [74] was used to evaluate the classification accuracy. As the name suggests, leave-one-out cross-validation uses the data from one subject as the testing data, and the data from the remaining subjects as the training data. This strategy is repeated to classify every subject.

Another analysis was performed in order to investigate the power of the proposed method for estimating the severity of OSA. The classification method used in this analysis was similar to the one described before for screening analysis. The only difference was that in this analysis we considered three classes: non-OSA ($AHI < 5$), moderate OSA ($10 \leq AHI < 25$), and severe OSA ($AHI \geq 30$). This analysis started with the same features that had passed the t-test from the previous analysis. In other words, the statistical t-test analysis was not repeated; instead, the features from the previous analysis that was based on discriminability for OSA screening were used. However, selection of the best subsets of features was performed in order to maximize the classification accuracy for the three-class problem. Furthermore, instead of the exhaustive search method, we adopted the forward floating search [77]. This was partly because the three-class classification used in this analysis was computationally more expensive than the two-class classification used in the previous analysis, making an exhaustive search of all feature combinations even more difficult. Moreover, as it will be explained in the next chapter, in classification for OSA screening, optimal number of features was 7; adding more features did not improve the classification performance on the testing data set. However, for the three-class classification used for OSA severity estimation, the classification accuracy continued to improve as the size of feature set was increased beyond 10. Therefore, it was necessary to adopt a more sophisticated approach than the exhaustive search.

Floating search methods are modifications of sequential forward or backward search methods. Sequential search methods suffer from the so-called nesting effect [74]. For example, in sequential backward method, once a feature is discarded, it will not be included again in a later step. Floating search techniques remove this constraint so that the decision to include or exclude a specific feature can be changed in future steps. In this study, a forward floating search method was used.

Assuming that the total number of available features is n, the goal of floating search method is to find the subsets of size k for $k = 1, 2, \dots m$ that optimize some cost function C. In this study we defined C to be the product of the classification accuracy on the training data set for the three classes involved. In other words:

$$C = (1 - e_{non-OSA}) * (1 - e_{moderate-OSA}) * (1 - e_{severe-OSA}), \quad (2.38)$$

where $e_{non-OSA}$, $e_{moderate-OSA}$, and $e_{severe-OSA}$ represent the classification error on the training data set for these three classes. Then, the floating search method tries to maximize this cost function. Suppose that $X_k = \{x_1, x_2, \dots, x_k\}$ is the best subset of size k. We use Y_{n-k} to denote the set of remaining features not present in X_k. Moreover, as the algorithm is making progress, it keeps a record of all the lower-dimension subsets of best features and their associated cost function. The algorithm is presented as a pseudo-code below. The basic idea is that, having found subsets of size $1, 2, \dots, k$, the best subset of size $k + 1$, X_{k+1}, is found by adding one feature from Y_{n-k}. Then, we return to the lower-dimension subset, and examine whether exclusion of any of the previously-selected features will improve the cost function. The answer to this question will determine if further backward search is performed or not. Details are provided in the pseudo-code.

Pseudo-code for forward floating search [74]

Step I. Inclusion- Choose the new feature to add: $x_{k+1} = \text{argmax}_{y \in Y_{n-k}} C(X_k, y)$. In other words, choose the feature that results in the largest improvement in the cost function. Set $X_{k+1} = \{X_k, x_{k+1}\}$.

Step II. Test

1. Find the feature that, if removed, has the least effect on the cost function: $x_r = \text{argmax}_{y \in X_{k+1}} C\,(X_{k+1} - \{y\})$.
2. If $r = k + 1$, change $k = k + 1$ and go to Step I.
3. If $r \neq k + 1$ and $C(X_{k+1} - \{x_r\}) < C(X_k)$, go to Step I.
4. If $k = 2$, set $X_k = X_{k+1} - \{x_r\}$ and $C(X_k) = C(X_{k+1} - \{x_r\})$, go to Step I.

Step III. Exclusion

1. Remove x_r: $X_k{}' = X_{k+1} - \{x_r\}$
2. Find the feature that, when removed from this new set, has the least effect on the cost function; i.e. $x_s = \text{argmax}_{y \in X_k{}'} C\,(X_k{}' - \{y\})$.
3. If $C(X_k{}' - \{x_s\}) < C(X_{k-1})$, then $X_k = X_k{}'$; go to Step I.
4. Set $X_{k-1}{}' = X_k{}' - \{x_r\}$ and $k = k + 1$.
5. If $k = 2$, set $X_k = X_k{}'$ and $C(X_k) = C(X_k{}')$; go to Step I.
6. Go to Step III.; part 1.

The algorithm was initialized by finding the best single feature, X_1, and best combination of two features, X_2, through an exhaustive search. The algorithm was terminated after finding the best feature subsets of size up to 31.

Table 2.3 shows details of the number of subjects used in different steps in this analysis. For training, data from 20 subjects with $AHI < 5$, 20 subjects with $10 \leq AHI < 25$, and 15 subjects with $AHI \geq 30$ were used. The selected best feature subsets were then used to classify the rest of the subjects. First, the subjects in the same three AHI ranges were classified. This included 59 subjects with $AHI < 5$, 25 subjects with $10 \leq AHI < 25$, and 14 subjects with $AHI \geq 30$. Then, the subjects within the AHI gaps, i.e. $5 \leq AHI < 10$ and $25 \leq AHI < 30$ were classified. Similar to the classification for OSA screening, all classifications were carried out using a leave-one-out cross-validation strategy.

Table 2.3. Breakdown of the number of subjects in classification for OSA severity estimation.

Range of AHI	Total no. of subjects	No. of subjects in training step	No. of subjects in the first testing step	No. of subjects in the second testing step
$AHI < 5$	79	20	59	-
$5 \leq AHI < 10$	26	-	-	26
$10 \leq AHI < 25$	45	20	25	-
$25 \leq AHI < 30$	10	-	-	10
$AHI \geq 30$	29	15	14	-

2.3.2 Classification Based on Risk Factors

In addition to and independent of the classification based on the features extracted from the breathing sounds, explained in the previous subsections, we classified the subjects based solely

on six risk factors. The risk factors included in this classification were: age, gender, BMI, Mallampati score, neck circumference, and smoking history. This was performed in order to compare the accuracy of the classification based on breathing sounds with the accuracy based solely on the risk factors.

In this approach, a relative risk was computed for each subject. The computation relied on the results of other studies that have estimated the contribution of different risk factors from large populations. The contributions of age, gender, BMI, and neck circumference were identified from a big study (n= 5615) on the major predictors of OSA in general population [35]. That study estimated the odds ratio for several risk factors and an AHI of 15 or higher by developing multiple linear regression models. We used the odds ratios for a model that included gender, BMI, age, and neck circumference as presented in [35]. To include the contribution of the Mallampati score and smoking history on relative risk, the results from two other studies were used [78, 51]. The following is a summary of the results from these three studies:

- Gender

  Compared to women, men have an odds ratio of 1.71 [35].

- Age

  Up to the age of 65, every 10-year increment in age will increase the odds ratio by a factor of 1.36. Beyond 65 years, the odds ratio will remain unchanged [35].

- BMI

  Every $5.3\ kg/m^2$ increment in BMI will increase the odds ratio by a factor that depends on age. For age in the ranges 34-45, 45-55, 55-65, 65-75, and 75-85 years, this factor is equal to 2.0, 1.8, 1.6, 1.5, and 1.3 respectively [35].

- Neck circumference

  Every $4.32\ cm$ increment in neck circumference will increase the odds ratio by a factor of 1.48 [35].

- Smoking

  The odds ratio for former smokers versus never-smokers is 1.86. Compared to never-smokers, the odds ratio for current smokers who smoke less than 20 cigarettes per day, between 20 and 40 cigarettes per day, and more than 40 cigarettes per day is 3.94, 3.25, and 6.74, respectively [51].

- Mallampati score

  Every 1-point increase in the Mallampati score increases the odds ratio approximately by a factor of 2 [78].

The relative risk for all subjects was initialized to 1. Then, for each of the risk factors, the relative risk of the corresponding subjects was multiplied by the odds ratio for that risk factor. For example, compared to women, men have 1.71 times the odds of having an AHI of 15 or greater [35]. Therefore, the relative risk for male subjects is multiplied by 1.71. This is not an exact approach, because the odds ratio is not a simple ratio of probabilities, as shown in the following equation:

$$odds\ ratio = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \tag{2.39}$$

Our approach disregards the $(1-p)$ terms because there is no closed-form equation that relates odds ratio to the ratio of the probabilities p_1/p_2. This does not create a large error because the probabilities in here are much smaller than 0.50. Furthermore, for some of the

subjects the information on risk factors was incomplete. For example, neck circumference had not been measured for some of the subjects. In order to be able to compute the relative risk for these subjects, the missing information was replaced by the average of the subjects in the same AHI group.

After computing the relative risk of the subjects, threshold(s) were identified to classify the subjects. For screening analysis, a single threshold was identified to separate the subjects into non-OSA and OSA groups. In other words, it is expected that the OSA subjects have a higher relative risk compared to non-OSA subjects. Using the same 70 subjects that were used for selecting the most characteristic features of the breathing sound signals, we found a threshold that could best separate these subjects. The same threshold was then used to classify the subjects in the training set. For the three-class classification required for OSA severity estimation, two thresholds were identified so that the subjects could be divided into three classes of non-OSA, moderate OSA, and severe OSA.

Chapter 3

Results

3.1 Classification Results for OSA Screening

3.1.1 Classification Based on Breathing Sound Features

As mentioned in Chapter 2, a total of 153 features were found to be statistically different between a randomly selected dataset including 20 non-OSA (AHI<10) and 15 OSA (AHI>20) subjects. Six of these features were discarded because they had an area under the ROC curve of smaller than 0.50. The best feature subsets were selected from among these 147 features. As mentioned in Chapter 2, subsets of 2, 3, and 4 features were selected by exhaustive search of all 147 features, whereas subsets of 5 to 10 features were selected by exhaustive search of a smaller set of features selected by the mRMR algorithm (Figure 2.4). For feature selection step, data from 35 non-OSA (AHI<10) and 35 OSA (AHI>20) subjects were used. This dataset included the 20 non-OSA and 15 OSA subjects that had been used in the initial statistical significance test.

None of the 147 features appeared consistently in all of the subsets. However, there were clear patterns in terms of the types of features that were selected more often. For example, most of the selected PSD features were selected from the sub-bands of 150-450 Hz, 450-600 Hz, 600-

1200 Hz, or from the entire frequency band of 100-2500 Hz. Regarding the bispectral features, the invariant features appeared most frequently in the best feature subsets. Moreover, the difference of the features between the nose and mouth breathing were selected quite frequently, whereas the difference of the features between the upright and supine breathing were selected less frequently. Appendix A provides details of the 7 features that together produced the lowest classification error on the training data set.

After identifying the best combinations of features, the prediction accuracy of each feature subset was evaluated in an unbiased manner respect to training data. For this purpose, first, the data from the 87 subjects in AHI<10 and AHI>20 groups that had not been included in feature selection stage was used. These subjects included 70 and 17 subjects in the AHI<10 and AHI>20 groups, respectively. A subset of 7 features produced the lowest prediction errors on these subjects. Specifically, this subset of 7 features, resulted in the accuracy, specificity, and sensitivity of 78%, 77%, and 82%, respectively, on the test dataset for classification of the two groups with AHI<10 and AHI>20 . Sensitivity, here, is the accuracy of classifiying the group with higher AHI.

$$specificity = \frac{TN}{TN+FP} \tag{3.1}$$

$$sensitivity = \frac{TP}{TP+FN} \tag{3.2}$$

$$overall\ accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.3}$$

where in the above equations, TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively.

Subsets of smaller or larger sizes resulted in lower accuracies. Therefore, we restricted our attention to the best subset of 7 features, the exact description of which is provided in Appendix A. It is worth mentioning that the accuracy, specificity, and sensitivity for classification of the training set using this best combination of 7 features were 81%, 79%, and 82%, respectively. The fact that the classification performance on the training and testing data sets were very close is a good indication of robustness of the proposed classification method. Table 3.1 shows the accuracy, specificity, and sensitivity for classification using the best combination of 7 features along with the results of classification based on risk factors for comparison.

Table 3.1. Summary of classification results for the test dataset including 70 subjects with AHI<10 and 17 subjects with AHI>20.

	Overall accuracy	Specificity	Sensitivity
Classification based on breathing sound features (best combination of seven features)	78%	77%	82%
Classification based on estimated relative OSA risk	68%	76%	35%

The best combination of 7 features was then used to classify the subjects in the middle group. This group included the subjects with an AHI value between 10 and 20. Since an AHI of 15 is considered as the border line between mild and moderate OSA, a classifier that could separate the subjects on the two sides of this border would be desirable. From the 32 subjects in the middle group, 16 had an AHI of below 15 and 16 had an AHI of above 15. From the 16 subjects with $10 \leq AHI < 15$, one of them did not have any of the feature values because of the noisy

signals, and therefore could not be classified. In this group, 10 (67%) were correctly classified as non-OSA and 5 (33%) were misclassified as OSA. From the 16 subjects with $15 < AHI \leq 20$, 7 of them (44%) were misclassified as non-OSA, and 9 of them (56%) were correctly classified as OSA.

Table 3.2 summarizes the result of classification of the middle group. In addition to the classification based on breathing sound features, the results of the classification based on risk factors is also provided in this table. It should be noted that in this table, classification accuracy for subjects with $10 \leq AHI < 15$ indicates those classified as non-OSA and classification accuracy for subjects with $15 < AHI \leq 20$ indicates those classified as OSA.

Table 3.2. Summary of classification results for the subjects in the middle group ($10 < AHI \leq 20$).

	Overall classification accuracy	Classification accuracy on subjects with $10 \leq AHI < 15$	Classification accuracy on subjects with $15 < AHI \leq 20$
Classification based on breathing sound features (best combination of seven features)	61%	67%	56%
Classification based on estimated relative OSA risk	41%	38%	44%

3.1.2 Classification Based on Risk Factors

Figure 3.1 shows a plot of the relative risk for the 70 subjects used in the training phase. As expected, on average, non-OSA subjects had a lower relative risk compared to OSA group. A threshold of 58.4 resulted in the highest accuracy in separating the OSA and non-OSA groups in the training set. The threshold is also shown in the figure. With this threshold, the overall accuracy, specificity, and sensitivity for classifying the training data are 68%, 71%, and 66%, respectively.

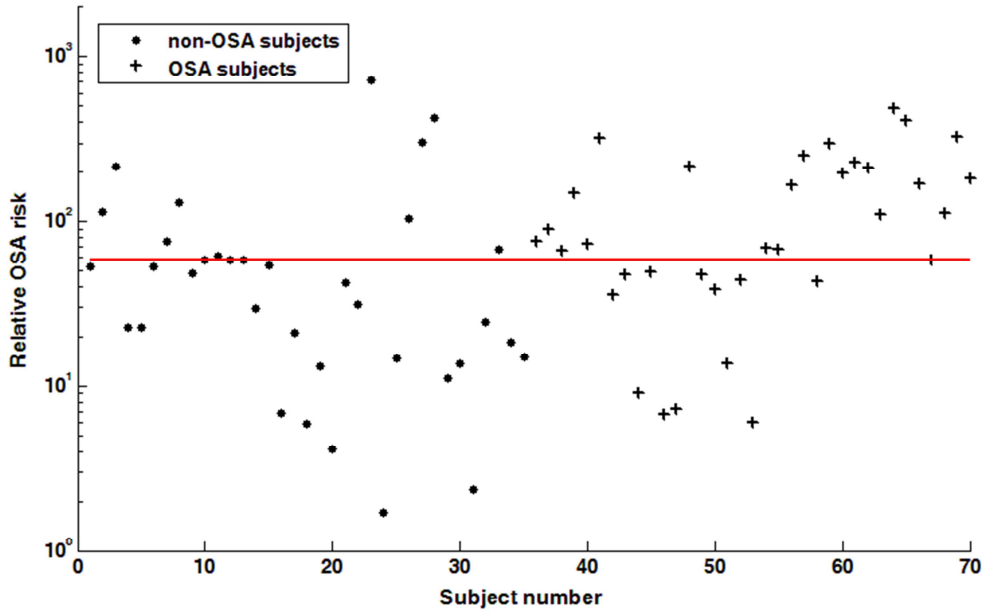


Figure 3.1. Relative OSA risk for the 70 subjects in the training dataset. The red horizontal line shows the threshold used for classification.

Applying the same threshold to classify the 87 subjects in the AHI<10 and AHI>20 groups that had not been used in the training phase, the classification accuracy, specificity, and sensitivity were 68%, 76%, and 35%, respectively, as shown in Figure 3.2 and Table 3.1. Considering the very different number of subjects in the two groups of the testing dataset (70 vs. 17), the overall

accuracy was biased towards the specificity. Nevertheless, it is clear that this method does not have a good sensitivity.

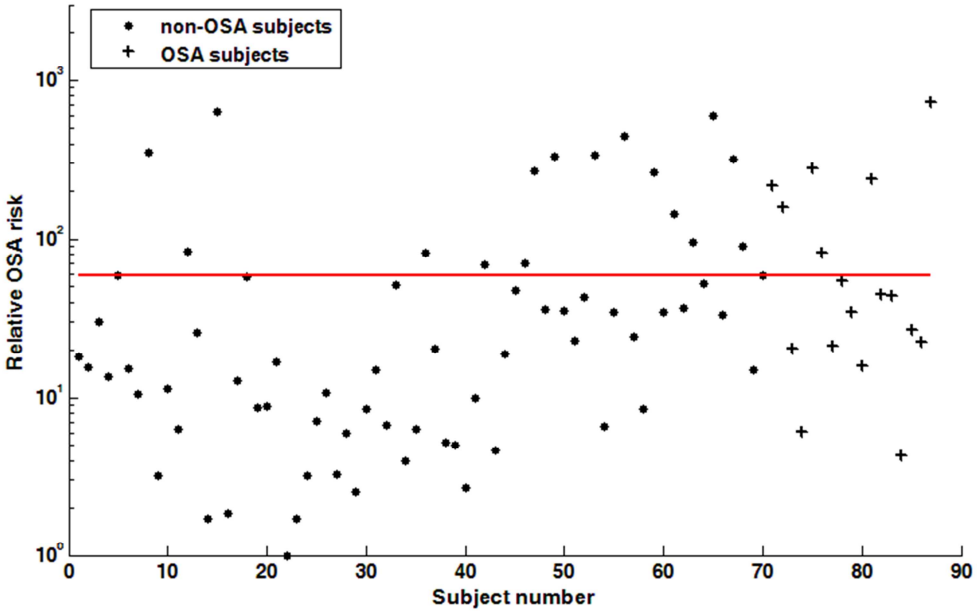


Figure 3.2. Relative OSA risk for the 87 subjects in the testing dataset. The red horizontal line shows the threshold used for classification.

The 32 subjects in the middle group were also classified based on relative risk and a threshold of 58.4. From the 16 subjects with $10 \leq AHI < 15$, 6 of them (38%) were correctly classified as non-OSA, and 10 of them (62%) were misclassified as OSA. From the 16 subjects with $15 < AHI \leq 20$, 9 of them (56%) were misclassified as non-OSA, and 7 of them (44%) were correctly classified as OSA. Therefore, it might be said that the overall accuracy was 41%. These results are shown in Table 3.2 along with the results of the same classification based on breathing sound features alone.

3.2 Classification Results for OSA Severity Estimation

3.2.1 Classification Based on Breathing Sounds

As mentioned in Chapter 2, estimation of OSA severity was performed by classifying the subjects into three groups of non-OSA ($AHI < 5$), moderate OSA ($10 \leq AHI < 25$), and severe OSA ($AHI \geq 30$). Selection of best set of features was performed through the use of the forward floating search algorithm on data from 20 subjects in the non-OSA group ($AHI < 5$), 20 subjects in the moderate OSA group ($10 \leq AHI < 25$), and 15 severe OSA subjects ($AHI \geq 30$). Best classification results on this training dataset were obtained with a feature set of size 22. Appendix A describes these features in detail. Table 3.3 shows the classification results on the training dataset.

Table 3.3. The results of the three-class classification based on breathing sound features on the training dataset.

		Assigned class		
		$AHI < 5$	$10 \leq AHI < 25$	$AHI \geq 30$
True class	$AHI < 5$	18	1	1
	$10 \leq AHI < 25$	5	13	2
	$AHI \geq 30$	1	0	14

The set of 22 features were then used to classify the subjects in the test dataset, which included 59 subjects in the non-OSA group ($AHI < 5$), 25 subjects in the moderate OSA group ($10 \leq AHI < 25$), and 14 severe OSA subjects ($AHI \geq 30$). Table 3.4 presents the classification results on the test dataset. The overall classification accuracy, defined as the percentage of

subjects that were classified in their true class, was 82% on the training dataset and 71% on the testing dataset. The classification accuracies for non-OSA ($AHI < 5$), moderate OSA ($10 \leq AHI < 25$), and severe OSA ($AHI \geq 30$) groups were 90%, 65%, and 93%, respectively, for training dataset, while these accuracies were 75%, 64%, and 71%, respectively, for testing dataset.

Table 3.4. The results of the three-class classification based on breathing sound features on the testing dataset.

		Assigned class		
		$AHI < 5$	$10 \leq AHI < 25$	$AHI \geq 30$
True class	$AHI < 5$	44	8	7
	$10 \leq AHI < 25$	6	16	3
	$AHI \geq 30$	3	1	10

Table 3.5 shows the classification results for the subjects, whose AHI score was not within any of the three classes defined for OSA severity estimation; this included 26 subjects with $5 \leq AHI < 10$ and 10 subjects with $25 \leq AHI < 30$. It is difficult to specify a classification accuracy value for these subjects. However, it is clear that the majority of subjects (i.e. 78% of them) were classified into one of the two neighboring classes.

3.2.2 Classification Based on Risk Factors

Figure 3.3 shows the estimated relative OSA risk for the 55 subjects in the training step. Thresholds that resulted in lowest classification errors were 59 and 148. Subjects with an estimated OSA risk of below 59 are classified as non-OSA, those with a risk above 148 are

classified as severe OSA, and subjects in between the two thresholds are classified as moderate-OSA. Table 3.6 shows the classification table for the training data using these thresholds. The classification accuracies for non-OSA, moderate-OSA, and severe-OSA groups were 75%, 55%, and 55% respectively.

Table 3.5. The results of the three-class classification based on breathing sound features for subjects with $5 \leq AHI < 10$ and $25 \leq AHI < 30$.

		Assigned class		
		$AHI < 5$	$10 \leq AHI < 25$	$AHI \geq 30$
True class	$5 \leq AHI < 10$	10	10	6
	$25 \leq AHI < 30$	2	2	6

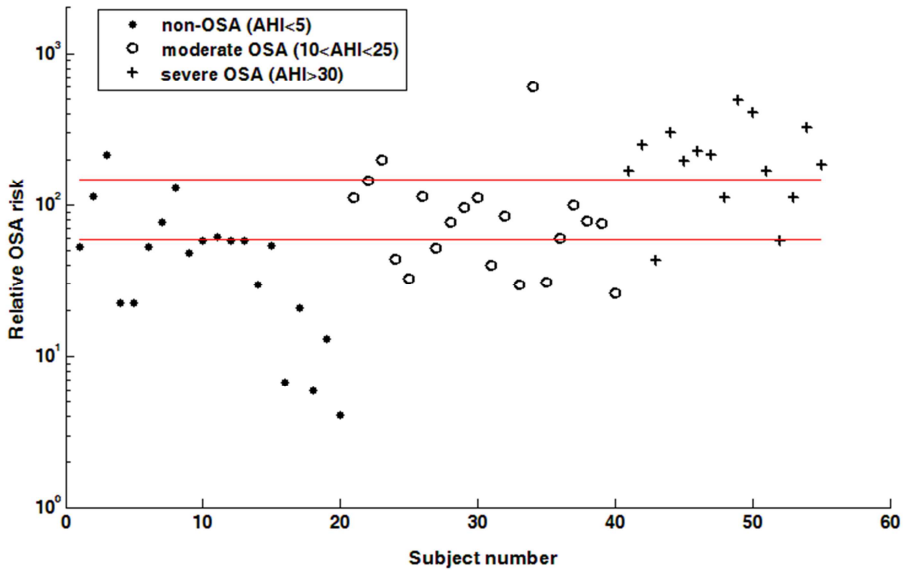


Figure 3.3. Relative OSA risk for the 55 subjects in the training dataset for estimation of OSA severity based on risk factors. The red horizontal lines show the thresholds used for classification.

Table 3.6. The results of the three-class classification based on estimated relative OSA risk on the training dataset.

		Assigned class		
		$AHI < 5$	$10 \leq AHI < 25$	$AHI \geq 30$
True class	$AHI < 5$	15	4	1
	$10 \leq AHI < 25$	7	11	2
	$AHI \geq 30$	2	2	11

The same thresholds were then used to classify subjects in the testing dataset. The classification accuracies for non-OSA, moderate-OSA, and severe-OSA groups were 83%, 28%, and 21% respectively. Detailed classification results for these subjects are presented in Table 3.7 and shown graphically in Figure 3.4.

Table 3.7. The results of the three-class classification based on estimated relative OSA risk on the testing dataset.

		Assigned class		
		$AHI < 5$	$10 \leq AHI < 25$	$AHI \geq 30$
True class	$AHI < 5$	49	4	6
	$10 \leq AHI < 25$	14	7	4
	$AHI \geq 30$	10	1	3

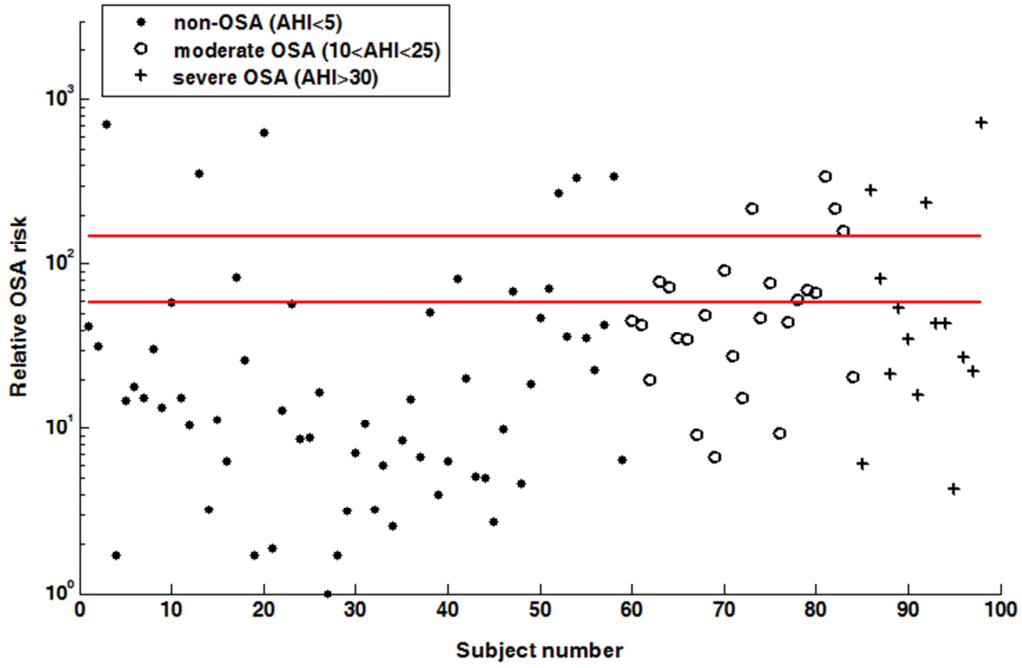


Figure 3.4. Relative OSA risk for the 98 subjects in the testing dataset for estimation of OSA severity based on risk factors. The red horizontal line shows the threshold used for classification.

Table 3.8 shows the classification results for subjects with $5 \leq AHI < 10$ and those with $25 \leq AHI < 30$ that were not included in Tables 3.6 and 3.7. This can be compared with Table 3.5 that shows a similar classification based on breathing sound features.

Table 3.8. The results of the three-class classification based on the estimated OSA risk for subjects with $5 \leq AHI < 10$ and $25 \leq AHI < 30$.

		Assigned class		
		$AHI < 5$	$10 \leq AHI < 25$	$AHI \geq 30$
True class	$5 \leq AHI < 10$	15	5	6
	$25 \leq AHI < 30$	6	2	2

Overall, the classification based on breathing sound is more accurate. In Table 3.8, 67% of subjects are classified into one of the two neighboring classes, compared to 78% for classification based on breathing sound features.

Chapter 4

Discussion

The subjects of this study were divided into three groups based on their AHI: non-OSA group (AHI<10; n=105), middle group (10≤AHI≤20; n=32), and the OSA group (AHI>20; n=52). To have reliable statistical tests, it is important to have enough data equally distributed in the groups. However, as noted above, our dataset was skewed toward non-OSA group, and the middle group did not have enough data to be used for both training and testing. Therefore, for training the algorithm, we selected equal numbers of subjects in the two lowest and highest AHI groups; thus, data from 35 of the subjects in non-OSA group and 35 of the subjects in the OSA group were used to select the best subsets of the features. The selected sets of features were first used to classify the remaining subjects in the non-OSA and OSA groups. The subjects in the middle group were classified using the same features with a threshold of AHI=15.

The classification results clearly show the potential of power spectral and bispectral features of the breathing sounds recorded during wakefulness for OSA screening. A total of 153 features were identified as statistically significantly different between non-OSA and OSA subjects ($p < 0.05$). Seventy five of these features had been extracted from PSD and the remaining 78 from the bispectrum of the signals. Although the main focus of this study was on classification

accuracy, the results of statistical analysis are also important because they indicate the significantly different features that represent underlying differences between the breathing sounds of non-OSA and OSA individuals. Forty six of the statistically significant features were actually the difference of the features between the nose and mouth breathing maneuvers and 30 of the features were the difference of the features between supine and upright postures, and these features were often among the best combinations of features for classification. This indicates that the breathing maneuvers of the protocol play an important role in achieving the goal of the study.

Among the PSD features, the signal power and spectral centroid were selected in the best feature subsets more frequently than the other features. These two features represent the intensity of the breathing sounds and the frequency of major PSD peaks in different frequency bands. Anatomical differences in the upper airway of non-OSA and OSA subjects lead to different airflow patterns that shift the frequency of major spectral peaks of breathing sounds. This shift in the frequency of major spectral peaks will, in turn, change the values of the signal power and spectral centroid in different frequency bands.

Among the bispectral features, the bispectral invariant feature and the area under the bispectrum in different parts of the non-redundant region were selected most often. Differences in the area under the bispectrum in different frequency bands between non-OSA and OSA subjects is indicative of the differences in nonlinear interactions, e.g. phase coupling, among the harmonic peaks of different frequencies. The precise interpretation of the bispectral invariant feature is difficult to be speculated without having a physical model of the respiratory system and investigating different scenarios of plausible changes due to OSA. Nevertheless, as described in section 2.2.2, this feature is the ratio of the imaginary and real parts of the area

under radial slices of the bispectrum in the non-redundant region. Clearly, the value of this feature is determined by the overall shape of bispectrum and the location of its major features. In general, the observed difference in bispectral features between non-OSA and OSA subjects may be due to more complex oscillation patterns on the airflow in the upper airway of OSA subjects. Studies have shown that under normal conditions, when upper airway is open, the airflow through upper airway is mostly a laminar flow [79, 80]. However, as the airway narrows, the flow can become turbulent [81]. We know that the Reynolds number, which is the most commonly used variable to predict the flow regime is inversely proportional to the diameter of the tube or conduit carrying the fluid [82]. Large Reynolds numbers are indicative of turbulent flow, which is dominated by chaotic eddies, vortices and other unstable flow patterns. In addition, for large Reynolds numbers, relative roughness of the surface of the tube becomes an increasingly important factor and contributes to the turbulence of the flow [82]. Therefore, a narrower upper airway in people with OSA may result in more complex flow patterns and breathing sounds in non-OSA people that are captured by bispectrum analysis.

In terms of classification accuracy based on breathing sound features, the most striking fact is that the training and prediction accuracies are very close, which is an indication of the robustness of the proposed method. The prediction accuracy is close to 80%, which is relatively high, considering the easy and fast nature of the recordings used in this method.

Since the proposed algorithm in this study claims to use breath sounds for OSA screening during wakefulness, we compared the performance of the proposed acoustic OSA screening with screening based on six risk factors (BMI, gender, age, smoking history, neck circumference, and Mallampati score). Our proposed method was based on estimating a relative OSA risk for each subject and using a threshold to separate non-OSA and OSA subjects. As shown in Tables 3.1 and

3.2, this method was considerably less accurate and less robust than screening based on breathing sound features. In addition to the approach based on estimated relative risk, a decision tree classifier was also developed using Iterative Dichotomiser 3 (ID3) algorithm [83] to classify the subjects based on risk factors. This algorithm uses the idea of information gain, defined as the reduction in the entropy. Building the tree from top down, at each node ID3 selects the feature that results in the highest information gain in the training dataset. This approach produces decision trees that achieve high accuracy on the training dataset, and are relatively short. Shorter decision trees are less likely to suffer from overfitting problems. The number of decision tree layers was set to 4. For continuous features such as BMI, the decision was based on the median of the feature value in the training set. The choice of the number of layers, as well as the design of the decision tree was based on achieving the lowest classification error rate on the same 70 subjects used for feature selection. The classification accuracies for the decision tree classifier were not better than those of the relative risk approach. For example, the classification accuracy, specificity, and sensitivity on the test dataset (70 subjects with AHI<10 and 17 subjects with AHI>20) was 66%, 70%, and 47%, respectively.

It is not obvious which of these techniques, i.e. estimating the relative contribution of each risk factor in an additive or multiplicative fashion versus a classifier approach, better models a physician's decision-making approach. It is likely that a classifier approach is intrinsically inadequate for this purpose. It is also possible that the low classification accuracies by the decision tree classifier in this study were mainly due to the small amount of data that was used to train the classifier. It is likely that a dataset that is proportionally large relative to the number of features could actually produce better results.

Nevertheless, the results obtained in this study are not considerably different from those reported in other studies that have used OSA risk factors and symptoms for OSA screening. For example, one study used snoring, BMI, age, and gender in a logistic regression model to predict OSA with $AHI \geq 10$ [84]. All four factors were significantly correlated with OSA. However, applying the model to 410 subjects, the sensitivity and specificity were 28% and 95%, respectively. The study also investigated the accuracy of subjective decision of a single clinician and found sensitivity and specificity of 52% and 70%, respectively. Generally, studies that have used OSA symptoms and OSA risk factors have reported higher accuracies. For example, one study used snoring, gasping at night, witnessed apneas, age, sex, and BMI to estimate an apnea risk index using multiple logistic regressions [85]. The estimated risk index was then used to classify 427 subjects. The sensitivity and specificity were 88% and 55%, respectively. A commonly used tool for OSA screening is the Berlin Questionnaire [86]; this questionnaire consists of three categories of questions: category 1 asks questions regarding presence, frequency, and loudness of snoring and witnessed apneas, category 2 questions ask about sleepiness and fatigue during the day as well as a question specifically about drowsy driving, while questions in category 3 ask about history of hypertension and BMI. If a subject fulfills the criteria in at least two of the three categories, they are considered to be at high risk of OSA. Many studies have investigated the accuracy of the Berlin Questionnaire, but the reported accuracy is highly varied. Most studies have reported sensitivity values between 35% and 67%, and specificity values between 55% and 64% [87-89]. However, there have been studies that have reported sensitivities around 80% but with specificity values between 39% and 58% [90, 91].

On the other hand, it is also clear that in classification based on estimated risk, one can decide to increase the sensitivity at the cost of specificity by simply lowering the threshold, or vice

versa. In general, a high sensitivity is more desirable because a misclassification of an OSA patient as non-OSA is more undesirable than misclassification of a non-OSA individual as an OSA patient. This is particularly true when a method is meant to be used for pre-screening. For example, if we choose a threshold of 20, instead of 58.4 (that maximized the overall classification accuracy on the training data set), the specificity and sensitivity on the training data will be 31% and 86% respectively, and specificity and sensitivity on testing dataset will be 51% and 82%.

Moreover, one may expect that a combination of classification based on breathing sound features and classification based on risk factors may provide a more powerful screening tool. One may speculate that the features computed from breathing sounds may contain some information that are already present in the risk factors. In other words, some of the differences observed in breathing sound features may be equivalently present in the risk factors, which are more easily measured. However, the wide difference between classification results based on breathing sound features and those based on risk factors can be interpreted as an indication that there is little overlap between these two sets of features. Nevertheless, since both of these approaches can be considered as easy and quick screening tools, it would be interesting to explore the possibility of combining them together for more accurate screening.

Because the classification based on breathing sound features was considerably more robust and more accurate, one approach to combine the two methods would be to use the classification based on breathing sounds as the primary classifier. Then, if this primary classifier cannot strongly associate a subject with any of the two classes, a classification based on risk factors is performed. We explored this approach by defining the strength of an assignment by the primary classifier as the number of the votes of the winning class. For example, considering the classifier

based on the best subset of 7 features, if the votes were divided 7 to 0, 6 to 1, or 5 to 2, it was considered a strong classification, whereas if the votes were divided 4 to 3, it was considered a weak classification. If a subject was weakly classified based on breathing sound features, we resorted to classification based on estimated OSA risk. Although this approach seems very reasonable and provided highly improved accuracy on the training dataset, the accuracy did not improve on the testing dataset.

A second approach would be to use both methods in parallel. In other words, each subject is classified by both of the classification methods and if a subject is classified as OSA by any of the two methods, the subject can be considered as high-risk and referred for more accurate tests such as PSG. Clearly, this approach is expected to increase the sensitivity at the cost of reduced specificity; as argued above, this is desirable for pre-screening. We explored this approach by using the classification based on best subset of 7 breathing sound features and classification based on estimated OSA risk using the same threshold computed for the training dataset (i.e. 21.1). The sensitivity in identifying the subjects in the OSA (AHI>20) group for both the training and testing datasets was 94%. The cost, as expected, was reduced specificity, which was 60% and 57% on training and testing datasets respectively.

As shown in Table 3.2, for subjects in the middle group ($10 \leq AHI < 20$), the classification accuracy was low for both classification based on breathing sound features and classification based on estimated OSA risk. This is not a surprising result considering that for classification based on breathing sound signals characteristic features had been selected based on their power in discriminating non-OSA and OSA groups, without any regard for the middle group. One should also bear in mind the continuous nature of AHI values in contrast to the artificially crisp nature of using a border-line AHI value, in this case $AHI = 15$, to separate the subjects into two

classes. For example, 5 of the subjects in the middle group had an AHI score of between 13 and 15, and 6 of the subjects had an AHI of between 15 and 17.6. It is not reasonable to expect that short recordings of breathing sounds during wakefulness be able to mark such fine differences.

Tables 3.3 to 3.8 summarize the results of three-class classification. As suggested before, this analysis can be thought of as an effort to estimate the severity of OSA. Similar to the analysis for OSA screening, breathing sound features resulted in a more accurate classification than the method based on estimated OSA risk. The classification accuracy on the testing dataset was 71%, which is almost 10% lower than the two-class classification. This is not surprising because with increasing number of possible classes the accuracy in assigning the true class should normally decrease. It is also worth noting that the classification accuracy for the moderate-OSA class (i.e. $10 \leq AHI < 25$) was considerably lower than the accuracy for non-OSA and severe-OSA classes. This may be partly due to the fact that the initial t-test analysis to identify statistically different features used the data from subjects in the two extreme AHI groups, i.e. $AHI < 5$ and $AHI \geq 30$.

The three-class classification based on estimated OSA risk was considerably less accurate. It seems that risk factors used in this study are not very promising for estimation of severity of OSA. However, as mentioned in section 2.3.2, relative OSA risk was estimated based on odds ratios that had been found in other studies with a threshold of $AHI = 15$. Therefore, it is reasonable to expect that using odds ratios to estimate the OSA risk, as done in this study, should produce better results when dividing the subjects into two groups rather than into three groups. A thorough search in the relevant databases did not find any previously-published studies to classify subjects into more than two classes using OSA risk factors or symptoms.

It is also not obvious how one can effectively combine the information in breathing sound features with the information provided by the risk factors for more accurate OSA severity estimation. One approach may be to use the estimated OSA risk to select the subjects that are at very low or very high OSA risk. In other words, a very low threshold for OSA risk can be specified below which a subject is classified as non-OSA and a very high threshold above which the subject is classified as severe OSA. The subjects with an estimated OSA risk between these two thresholds are then classified using breathing sound features. We investigated this approach with various values for the low and high thresholds, but could not achieve any improvements beyond the classification results based on breathing sound features alone.

Overall, the results of this study are very promising. The methods suggested in this study provide an easy and powerful OSA screening method. As mentioned in Chapter 1, the gold standard for OSA diagnosis remains the overnight PSG. However, in view of the limited resources such as recording beds and skilled technicians, high cost, and long waiting lists, many researchers have explored the potential of OSA symptoms and risk factors for identifying higher-risk individuals. Screening devices in the form of single or multi-channel monitoring systems have also become common and represent an alternative OSA screening method. The reported accuracy of single-channel systems has been very varied with sensitivity and specificity values ranging from 41% to above 90% [92-95]. In general, the accuracy of multichannel monitoring systems is about 80%, with more channels improving the accuracy at the cost of adding cost and complexity [96, 20]. Even though many such devices are commercially available and can be used at the patients' home without the need for expert supervision [97, 98], their role and use is still controversial [20]. As explained previously in this chapter, screening based on OSA risk factors and symptoms has also been explored quite extensively, with mixed results even for the widely used Berlin Questionnaire.

Being able to screen patients for OSA before PSG is an important issue because the full-night PSG is costly and time-consuming, and there is usually a long waiting list for this test. In addition, many of the patients referred for full-night PSG are non-OSA. For example, all of the 189 subjects in this study had been referred for PSG by a physician, but 79 of them (42%) had an AHI of less than 5. Therefore, powerful screening tools can be very helpful for more efficient use of limited facility and resources such as PSG by identifying individuals who are at higher risk. The screening method proposed in this study relies on a combination of breathing sound features and major OSA risk factors. Although it is a simple and fast technique, its accuracy is very good compared to other OSA screening methods and devices currently used as alternatives to PSG. Therefore this study is a major step in development of fast and reliable OSA screening methods that can lead to efficient use of expensive and limited resources such as sleep laboratories and their personnel and reduction in the number of undiagnosed OSA patients.

Chapter 5

Conclusion and Recommendations for Future Work

5.1 Conclusion

This study demonstrated the potential of second and third-order spectral features of breathing sounds for OSA screening and OSA severity estimation. The proposed classification technique based solely on breathing sound features was able to achieve acceptable accuracy and very high robustness in classifying subjects with AHI>20 from subjects with AHI<10. To the best of our knowledge there is no previously published work on OSA diagnosis during wakefulness with more than 52 subjects. Therefore, this study can be regarded as a landmark because it increases the hopes for fast and inexpensive OSA screening during wakefulness.

It is unlikely that techniques such as the one proposed in this study will completely replace full-night PSG. However, they may provide very powerful screening tools for identifying patients that are more likely to suffer from OSA. Therefore, one of the potential applications of this technology is for identifying the patients at higher risk of OSA for PSG referral. The proposed screening method based on breathing sound features outperformed the suggested screening method based on OSA risk factors by a wide margin. The risk factors included in this study were

among the most significant OSA risk factors. In addition to these risk factors, a physician may use OSA symptoms such as witnessed snoring or apneas (by a bed partner) when referring a patient for PSG. However, this information is not always available or accurate. Therefore, it is reasonable to claim that the results of this study demonstrate the potential of the proposed screening method based on breathing sounds as a reliable alternative screening tool.

Using the classification based on breathing sounds in parallel with a classification based on risk factors resulted in very high sensitivity at the cost of very low specificity. With this approach, only one of the 17 subjects with AHI>20 was misclassified in the testing dataset. Recording of breathing sounds as suggested in this study can be performed in a very short time in a physician's office. Risk factors can also be measured by an interview and simple measurements. Therefore, the great advantage of the methods suggested in this study is the ease and speed. Even though parallel classification using breathing sounds and risk factors has low specificity, it should be noted that all of subjects in this study had been referred for full-night PSG by a physician. Using the methods suggested in this study, 54 of the subjects with AHI<10 (i.e. 51% of the 105 subjects in this class) had been classified as non-OSA by both classification using breathing sounds and classification using risk factors. This is a significant saving in expensive and time-consuming resources such as sleep laboratories that could be used for diagnosis of OSA patients.

Estimation of OSA severity based on breathing sounds is a more challenging task. This study achieved a maximum classification accuracy of 71% on testing dataset when classifying the subjects into three different classes of OSA severity. The severity estimation based on estimated OSA risk was less accurate, probably due to the nature of the OSA risk estimation method used in this study.

Overall, the results of this study showed the potential of breathing sounds recorded during wakefulness for OSA screening. The proposed classification proved much more accurate than classification based on risk factors alone. It was possible to combine the two methods to achieve a higher sensitivity at the cost of specificity. Considering the fact that the subjects in this study had all been referred for full-night PSG, the proposed method may lead to significant reduction in the waiting time for comprehensive tests such as PSG, reduction in the number of undiagnosed OSA patients, and substantial decrease in the cost of undiagnosed OSA for the healthcare system.

5.2 Recommendations for future work

I. This study used AHI from the PSG test as the gold standard for severity of OSA. Even though no other metric has proven to be better than AHI [33], AHI does not consider some very important processes such as the degree of oxygen desaturation or the total number of arousals from sleep [33]. Therefore, it would be valuable to compare the results of the classification methods developed in this study with other OSA metrics.

II. This study focused only on features extracted from the estimated PSD and bispectrum of the breathing sound signals. Bispectral features were frequently selected among the best feature combinations. It is likely that other higher-order polyspectra such as the trispectrum would reveal additional information in the breathing sounds that are hidden to PSD and bispectrum. It is also likely that estimation of the bispectrum at a higher frequency resolution, than $5\ Hz$ used in this study, will improve the results further. We were unable to explore these possibilities due to our limited computational resources.

III. The greatest advantage of the proposed screening method proposed in this study is the ease and speed with which the required information can be collected. Breathing sounds

and the information about risk factors can be collected very easily. In addition, OSA symptoms such as daytime sleepiness, snoring, and witnessed apneas are also strong predictors of OSA. Future studies can explore whether including the information about OSA symptoms can further improve the accuracy of the methods proposed in this thesis.

IV. The proposed OSA screening method in this study is based on classifying each subject twice, once based on breathing sound features and once based on the estimated OSA risk. An alternative approach would be to match the subjects based on the risk factors and then perform the classification based on breathing sounds. This approach will require a large databank of breathing sounds so that a new subject could be matched with an adequate number of subjects within the same age, gender, and BMI group. The classification of the new subject as non-OSA or OSA will be performed by comparing his or her breathing sound features with those of the matched subjects.

V. The results of this study show that the change in PSD and bispectral features of the breathing sounds between nose and mouth breathing and between upright and supine postures are significantly different between non-OSA and OSA subjects. In addition to breathing maneuver (nose or mouth breathing) and posture, it is likely that airflow rate can also be an influential factor. This study only looked at the breathing sound features at deep breathing with maximum airflow rate. Future studies can look at the change in PSD and bispectral features between shallow or normal breathing and deep breathing. Characteristics of a flow such as flow regime, i.e. laminar or turbulent, strongly depend on the flow velocity [82]. Therefore, it is likely that airflow rate through the upper airway will significantly alter the characteristics of breathing sounds and this change is probably different between non-OSA and OSA individuals.

VI. As described in Chapter 2, in this study 4 breathing sound signals were recorded from each subject and each signal contained 5 inspiration/expiration cycles, resulting in 40 inspiration and expiration cycles. Estimation of PSD and bispectrum using the techniques described in this thesis for one subject takes approximately 12 minutes using Matlab software (Mathworks, Natick, MA) on a PC with a 1.73 GHz processor. It would be valuable to explore possible ways to decrease this time. For example, if the collected signal is free from noise and artefacts, it may not be necessary to compute the features from all five respiratory cycles in each signal. Also, it may be possible to reduce the frequency resolution used in this study for estimation of PSD ($1\ Hz$) and bispectrum ($5\ Hz$) without any loss of classification accuracy.

VII. Biological signals, such as the breathing sounds analyzed in this study, are characterized by their huge variability. A full understanding of the underlying variability in such signals requires a very large set of measurements. In this study, data from 189 subjects were used, 70 of which were used for statistical significance test and for selection of the best feature subsets. A complete understanding of the variability of the spectral and bispectral features of breathing sounds requires a much larger dataset. This is particularly true if the breathing sounds are to be used for estimation of OSA severity. Therefore, future studies should consider increasing the number of subjects.

Bibliography

[1]. National Sleep Foundation, www.sleepfoundation.org, 2012.

[2]. P. Peppard, T. Young, M. Palta, J. Skatrud, “Prospective study of the association between sleep disordered breathing and hypertension,” N Engl J Med, vol. 342, pp. 1378–84, 2000.

[3]. E. Shahar, C. Whitney, S. Redline, E. Lee, A. Newman, F. Javier Nieto, G. O'Connor, L. Boland, J. Schwartz, and J. Samet, “Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study,” Am J Respir Crit Care Med, vol. 163, pp. 19-25, 2001.

[4]. E. Lindberg, N. Carter, T. Gislason, C. Janson, “Role of snoring and daytime sleepiness in occupational accidents,” Am J Respir Crit Care Med, vol. 164, pp. 2031–35, 2001.

[5]. J. Teran-Santos, A. Jimenez-Gomez, J. Cordero-Guevara, and Cooperative Group Burgos-Santander, “The association between sleep apnoea and the risk of traffic accidents,” N Engl J Med, vol. 340, pp. 847-51, 1999.

[6]. American Academy of Sleep Medicine, “International classification of sleep disorders, 2nd edition: diagnostic and coding manual,” Westchester, IL: American Academy of Sleep Medicine, 2005.

[7]. T. Morgenthaler, V. Kagramanov, V. Hanak, and P. Decker, “Complex sleep apnea syndrome: is it a unique clinical syndrome?” Sleep, vol. 29, pp. 1203-9, 2006

[8]. T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, “The occurrence of sleep-disordered breathing among middle-aged adults,” N Engl J Med, vol. 32, pp. 1230–35, 1993.

[9]. J. Durán, S. Esnaola, R. Rubio, and A. Iztueta, “Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr.,” Am J Respir Crit Care Med, vol. 163, pp. 685-9, 2001.

[10]. E. Bixler, A. Vgontzas, H. Lin, T. Ten Have, J. Rein, A. Vela-Bueno, and A. Kales, “Prevalence of sleep-disordered breathing in women: effects of gender,” Am J Respir Crit Care Med, vol. 163, pp. 608-13, 2001.

[11]. 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. 9, pp. 705-6, 1997.

[12]. V. Kapur, D. Blough, R. Sandblom, R. Hert, J. de Maine, S. Sullivan, B. Psaty, “The medical cost of undiagnosed sleep apnea,” Sleep, vol. 22, pp. 749-55, 1999.

[13]. T. M. Davidson, “The great leap forward: The anatomic basis for the acquisition of speech and obstructive sleep apnea,” Sleep Med., vol. 4, no. 3, pp. 185–194, 2003.

[14]. J. Remmers, W. deGroot, E. Sauerland, and A. Anch, “Pathogenesis of upper airway occlusion during sleep,” J Appl Physiol, vol. 44, pp. 931–38, 1978.

[15]. I. Kobayashi, A. Perry, Rhymer J, B. Wuyam, P. Hughes, K. Murphy, J. Innes, J. McIvor, A. Cheesman, and A. Guz, “Inspiratory coactivation of the genioglossus enlarges retroglossal space in laryngectomized humans,” J Appl Physiol, vol. 80, pp. 1595–604, 1996.

[16]. S. Isono, J. Remmers, A. Tanaka, Y. Sho, J. Sato, and T. Nishino, “Anatomy of pharynx in patients with obstructive sleep apnoea and in normal subjects,” J Appl Physiol, vol. 82, pp. 1319–26, 1997.

[17]. M. Younes, “Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders,” J. Appl. Physiol, vol. 105, pp. 1389-1405, 2008.

[18]. T. Akahoshi, D. White, J. Edwards, J. Beauregard, and S. Shea, “Phasic mechanoreceptor stimuli can induce phasic activation of upper airway muscles in humans,” J Physiol, vol. 531, pp. 677–91, 2001.

[19]. W. Mezzanotte, D. Tangel, and D. White, “Waking genioglossal electromyogram in sleep apnoea patients versus normal controls (a neuromuscular compensatory mechanism),” J Clin Invest, vol. 89, pp. 1571–79, 1992.

[20]. A. Malhotra, and D. White, “Obstructive sleep apnoea,” The Lancet, vol. 360, pp. 237-245, 2002.

[21]. J. Wheatley, W. Mezzanotte, D. Tangel, and D. White, “Influence of sleep on genioglossus muscle activation by negative pressure in normal men,” Am Rev Respir Dis, vol. 148, pp. 597–605, 1993.

[22]. W. Van de Graaff, “Thoracic traction on the trachea: mechanisms and magnitude,” J Appl Physiol, vol. 70, pp. 1328–63, 1991.

[23]. M. Younes, M. Ostrowski, W. Thompson, C. Leslie, and W. Shewchuk, “Chemical control stability in patients with obstructive sleep apnoea,” Am J Respir Crit Care Med, vol. 163, pp. 1181–90, 2001.

[24]. D. White, T. Gibb, J. Wall, and P. Westbrook, “Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home,” Sleep, vol. 18, pp. 115-126, 1995.

[25]. E. Chiner, J. Signes-Costa, J. Arriero, J. Marco, I. Fuentes, A. Sergado, “Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies?,” Thorax, vol. 54, pp. 968-71, 1999.

[26]. C. Hsu, and P. Shih, “A novel sleep apnea detection system in electroencephalogram using frequency variation,” Expert Systems with Applications, vol. 38, pp. 6014–6024, 2011.

[27]. J. Salisbury, and Y. Sun, “Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique,” Med Eng Physc, vol. 9, pp. 336–343, 2007.

[28]. P. Caseiroa, R. Fonseca-Pintoa, and A. Andrade, “Screening of obstructive sleep apnea using Hilbert–Huang decomposition of oronasal airway pressure recordings,” Med Eng Physc, vol. 32, pp. 561–568, 2010.

[29]. E. Goldshtein, A. Tarasiuk, and Y. Zigel, “Automatic detection of obstructive sleep apnea using speech signals,” IEEE Trans Biom Eng, vol. 58, pp. 1373-1382, 2011.

[30]. H. Pasterkamp, J. Schäfer, and G. Wodicka, “Posture-dependent change of tracheal sounds at standardized flows in patients with obstructive sleep apnea,” Chest, vol. 110, pp. 1493-8, 1996.

[31]. A. Yadollahi, E. Giannouli, and Z. Moussavi, "Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals", J. Med. Biol. Eng. Comput., vol. 48, No 11, pp. 1087-97, 2010.

[32]. A. Montazeri, E. Giannouli, and Z. Moussavi, “Assessment of obstructive sleep apnea and its severity during wakefulness,” J. Annals on Biomed. Eng., vol. 4, No.4, pp. 916-24, 2012.

[33]. S. Caples, A. Gami, V. Somers, “Obstructive sleep apnea,” Ann Intern Med, vol. 142, pp. 187-197, 2005.

[34]. T. Young, P. Peppard, and D. Gottlieb, “The epidemiology of obstructive sleep apnoea: a population health perspective,” Am J Respir Crit Care Med, vol. 165, pp. 1217–39, 2002.

[35]. T. Young, E. Shahar, F. J. Nieto, S. Redline, A. B. Newman, D. J. Gottlieb, J. A. Walsleben, L. Finn, P. Enright, and J. M. Samet, “Predictors of sleep-disordered breathing in community-dwelling adults: the Sleep Heart Health Study,” Arch Intern Med, vol. 162, pp. 893-900, 2002.

[36]. A. Krystal, J. Edinger, W. Wohlgemuth, and G. Marsh, “Sleep in peri-menopausal and post-menopausal women,” Sleep Med Rev, vol. 2, pp. 243–253, 1998.

[37]. I. Waldron, “What do we know about causes of sex differences in mortality? A review of the literature,” In: P. Conrad P, and R. Kern R, editors, “The sociology of health and illness: critical perspectives,” NY: St. Martin’s Press; 1994.

[38]. K. Strohl, S. Redline, “Recognition of obstructive sleep apnea,” Am J Respir Crit Care Med, vol. 154, pp. 274–289, 1996.

[39]. J.R. Stradling and J.H. Crosby, “Predictors and prevalence of obstructive sleep apnoea and snoring in 1001 middle aged men,” Thorax, vol. 46, pp. 85–90, 1991.

[40]. P.L. Enright, A.B. Newman, P.W. Wahl, T.A. Manolio, E.F. Haponik, and P.J. Boyle, “Prevalence and correlates of snoring and observed apneas in 5,201 older adults,” Sleep, vol. 19, pp. 531–538, 1996.

[41]. V.A. Barvaux, G. Aubert, and D.O. Rodenstein, “Weight loss as a treatment for obstructive sleep apnoea,” Sleep Med Rev, vol. 4, pp. 435–452, 2000.

[42]. R.J. Strobel, and R.C. Rosen, “Obesity and weight loss in obstructive sleep apnea: a critical review,” Sleep, vol. 19, pp. 104–115, 1996.

[43]. R.J. Davies and J.R. Stradling, “The relationship between neck circumference, radiographic pharyngeal anatomy, and the obstructive sleep apnoea syndrome, Eur Respir J, vol. 3, pp. 509–514, 1990.

[44]. P.D. Levinson, S.T. McGarvey, C.C. Carlisle, S.E. Eveloff, P.N. Herbert, and R.P. Millman, “Adiposity and cardiovascular risk factors in men with obstructive sleep apnea,” Chest, vol. 103, pp. 1336–1342, 1993.

[45]. E. Shinohara, S. Kihara, S. Yamashita, M. Yamane, M. Nishida, T. Arai, K. Kotani, T. Nakamura, K. Takemura, and Y. Matsuzawa, “Visceral fat accumulation as an important risk factor for obstructive sleep apnoea syndrome in obese subjects,” J Intern Med, vol. 241, pp. 11–18, 1997.

[46]. S.Ancoli-Israel, D. Kripke, M. Klauber, W. Mason, R. Fell, and O. Kaplan, “Sleep-disordered breathing in community-dwelling elderly,” Sleep, vol. 14, pp. 486–495, 1991.

[47]. C.J. Durán, “Prevalence of obstructive sleep apnea–hypopnea and related clinical features in the elderly: a population-based study in the general population aged 71–100,” World Conference 2001 Sleep Odyssey, Montevideo, Uruguay, October 21–26, 2001.

[48]. P. Jennum, H.O. Hein, P. Suadicani, and F. Gyntelberg, “Cardiovascular risk factors in snorers: a cross-sectional study of 3,323 men aged 54 to 74 years. The Copenhagen Male Study,” Chest, vol. 102, pp. 1371–1376, 1992.

[49]. J.W. Bloom, W.T. Kaltenborn, and S.F. Quan, “Risk factors in a general population for snoring: importance of cigarette smoking and obesity,” Chest, vol. 93, pp. 678–683, 1988.

[50]. A.B. Newman, F.J. Nieto, U. Guidry, B.K. Lind, S. Redline, E. Shahar, T.G. Pickering, and S.F. Quan, “Relation of sleep-disordered breathing to cardiovascular disease risk factors: The Sleep Heart Health Study,” Am J Epidemiol, vol. 154, pp. 50–59, 2001.

[51]. D. Wetter, T. Young, T. Bidwell, M. Badr, and M. Palta, “Smoking as a risk factor for sleep-disordered breathing,” Arch Intern Med, vol. 154, pp. 2219-24, 1994.

[52]. R.W. Robinson, D.P. White, and C.W. Zwillich, “Moderate alcohol ingestion increases upper airway resistance in normal subjects,” Am Rev Respir Dis, vol. 132, pp. 1238–1241, 1985.

[53]. W. Tsutsumi, S. Miyazaki, Y. Itasaka, and K. Togawa, “Influence of alcohol on respiratory disturbance during sleep,” Psychiatry Clin Neurosci, vol. 54, pp. 332–333, 2000.

[54]. M.M. Mitler, A. Dawson, S.J. Henriksen, M. Sobers, and F.E. Bloom, “Bedtime ethanol increases resistance of upper airways and produces sleep apneas in asymptomatic snorers,” Alcohol Clin Exp Res, vol. 12, pp. 801–805, 1988.

[55]. L. Scrima, P.G. Hartman, and F.C. Hiller, “Effect of three alcohol doses on breathing during sleep in 30–49 year old non-obese snorers and non-snorers,” Alcohol Clin Exp Res, vol. 13, pp. 420–427, 1989.

[56]. S. Redline, P.V. Tishler, M.G. Hans, T.D. Tosteson, K.P. Strohl, and K. Spry, “Racial differences in sleep-disordered breathing in African-Americans,” Am J Respir Crit Care Med, vol. 155, pp. 186–92, 1997.

[57]. T. Young, “Menopause, hormone replacement therapy, and sleep-disordered breathing: are we ready for the heat,” Am J Respir Crit Care Med, vol. 163, pp. 597–98, 2001.

[58]. S.R. Mallampati, S.P. Gatt, L.D. Gugino, S.P. Desai, B. Waraksa, D. Freiberger, and P.L. Liu, “A clinical sign to predict difficult tracheal intubation: a prospective study,” Can Anaesth Soc J, vol. 32, no. 4, pp. 429-34, 1985.

[59]. J.L. Benumof, “Management of the difficult adult airway; with special emphasis on awake tracheal intubation,” Anesthesiology, vol. 75, pp. 1087-110, 1991.

[60]. Nikias, C. L. and M. R. Raghuveer, “Bispectrum Estimation: A Digital Signal Processing Framework,” Proceedings IEEE, 75(7), pp. 86S891, July 1987.

[61]. Nikias, C.L. and J.M. Mendel, “Signal processing with higher-order spectra,” IEEE Signal Processing Magazine, Vol. 10, No 3, pp. 10-37, 1993.

[62]. M. J. Hinich, "Testing for Gaussianity and linearity of a stationary time series," J Time Series Analysis, vol. 3, pp. 169-76, 1982.

[63]. N. Gavriely, Y. Palti, and G. Alroy, "Spectral characteristics of normal breath sounds," J Appl Physiol, vol. 50, pp. 307–314, 1981.

[64]. A.V. Oppenheim, A.S. Willsky, S. Hamid, Signals and Systems, 2nd Edition, Prentice Hall, 1996.

[65]. D. FitzGerald and J Paulus, "Unpitched Percussion Transcription" in A. Klapuri and M. Davy, Editors, "Signal Processing Methods for Music Transcription," 1st Edition, Springer, 2006.

[66]. C. Rauscher, "Fundamentals of Spectrum Analysis," 1ST Edition, Rohde & Schwarz, 2001.

[67]. P. Stoica, and R. Moses, "Introduction to Spectral Analysis," Prentice Hall, Upper Saddle River, NJ, 1997.

[68]. X. Chen, X. Zhu, and D. Zhang, "A discriminant bispectrum feature for surface electromyogram signal classification," Med Eng Phys, vol. 32, pp. 126–135, 2010.

[69]. S. Yu, and M. Lee, "Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability," Comput Biol Med, vol. 42, pp. 816–825, 2012.

[70]. M. Mohebbi, and H. Ghassemian, "Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal," Comput Meth Prog Bio, vol. 105, pp. 40-49, 2012.

[71]. S. Zhou, J. Gan, F. Sepulveda, “Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface,” Inform Sciences, vol. 178, pp. 1629–1640, 2008.

[72]. V. Chandran, and S. Elgar, “Pattern recognition using invariants defined from higher order spectra-one dimensional inputs,” IEEE Transactions on Signal Processing, vol. 41, pp. 205-12, 1993.

[73]. H. Pasterkamp, S.S. Kraman, G.R. Wodicka, “Respiratory sounds. Advances beyond the stethoscope,” Am J Respir Crit Care Med, vol. 156, pp. 974-87, 1997.

[74]. S. Theodoridis, and K. Koutroumbas, “Pattern Recognition,” 4th Ed., Academic Press, 2008.

[75]. R. Duda, P. Hart, and D. Stork, “Pattern Classification,” 2nd Ed., Wiley-Interscience, 2000.

[76]. H. Peng, F. Long, and C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on pattern analysis and machine intelligence, pp. 1226{1238, 2005.

[77]. P. Pudil, J. Novovicova, and J. Kittler, “Floating search methods in feature selection,” Pattern Recognition Letters, vol. 15, pp. 1119–1125, 1994.

[78]. T. Nuckton, D. Glidden, W. Browner, D. Claman, “Physical examination: Mallampati score as an independent predictor of obstructive sleep apnea,” Sleep, vol. 29, 903-8, 2006.

[79]. Y. Huang, D.P. White, and A. Malhotra, “Use of computational modeling to predict responses to upper airway surgery in obstructive sleep apnea,” The Laryngoscope, vol. 117, pp. 648-53, 2007.

[80]. W. Vos, J. De Backer, A. Devolder, O. Vanderveken, S. Verhulst, R. Salgado, P. Germonpre, B. Partoens, F. Wuyts, P. Parizel, and W. De Backer, “Correlation between severity of sleep apnea and upper airway morphology based on advanced anatomical and functional imaging,” J Biomech, vol. 40, pp. 2207–13, 2007.

[81]. Y. Fan, L.K. Cheung, M.M. Chong, H.D. Chua, K.W. Chow, and C.H. Liu, “Computational fluid dynamics analysis on the upper airways of obstructive sleep apnea using patient - specific models,” IAENG Int J Compt S, vol. 38, pp. 401-408, 2011.

[82]. R.W. Fox, A.T. McDonald, P.J. Pritchard, Introduction to Fluid Mechanics, 6th Ed., New York City, NY: John Wiley & Sons, Inc., 2003.

[83]. B. Coppin, Artificial Intelligence Illuminated, Sudbury, MA: Jones and Bartlett Publishers, Inc., 2004.

[84]. S. Viner, J.P. Szalai, and V. Hoffstein, “Are the history and physical examination a good screening test of OSA?” Ann Intern Med, vol. 115, pp. 356-9, 1991.

[85]. G. Maislin, A.I. Pack, and N.B. Kribbs, “A survey screen for the prediction of apnea,” Sleep, vol. 18, pp. 158- 66, 1995.

[86]. N.C. Netzer, R.A. Stoohs, C.M. Netzer, K. Clark, and K.P. Strohl, “Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome,” Ann Intern Med, vol. 131, pp.485–91, 1999.

[87]. P.R. Srijithesh, G. Shukla, A. Srivastav, V. Goyal, S. Singh, and M Behari, “Validity of the Berlin Questionnaire in identifying obstructive sleep apnea syndrome when administered to the informants of stroke patients,” J Clin Neurosci, vol. 18, pp. 340–343, 2011.

[88]. S.A. Olivarez, B. Maheshwari, M. McCarthy, N. Zacharias, I. van den Veyver, L. Casturi, H. Sangi-Haghpeykar, and K. Aagaard-Tillery, “Prospective trial on obstructive sleep apnea in pregnancy and fetal heart rate monitoring,” Am J Obstet Gynecol, vol. 202, no. 6, pp. 552.e1-7, 2010.

[89]. P.R. Srijithesh, G. Shukla, A. Srivastav, V. Goyal, S. Singh, M. Behari, “Validity of the Berlin Questionnaire in identifying obstructive sleep apnea syndrome when administered to the informants of stroke patients,” J Clin Neurosci, vol. 18, no. 3, pp. 340-3, 2011.

[90]. M.J. Thurtell, B.B. Bruce, D.B. Rye, N.J. Newman, V. Biousse, “The Berlin questionnaire screens for obstructive sleep apnea in idiopathic intracranial hypertension,” J Neuroophthalmol, vol. 31, no. 4, pp. 316-9, 2011.

[91]. E. Sforza, F. Chouchou, V. Pichot, F. Herrmann, J.C. Barthélémy, F. Roche, “Is the Berlin questionnaire a useful tool to diagnose obstructive sleep apnea in the elderly?” Sleep Med, vol. 12, no. 2, pp. 142-6, 2011.

[92]. N. Wiltshire, A.H. Kendrick, and J.R. Catterall, “Home oximetry studies for diagnosis of sleep apnea/hypopnea syndrome: limitation of memory storage capabilities,” Chest, vol 120, pp. 384-9, 2001.

[93]. T. Shochat, N. Hadas, M. Kerkhofs, A. Herchuelz, T. Penzel, J.H. Peter, and P. Lavie, “The SleepStrip: an apnoea screener for the early detection of sleep apnoea syndrome. Eur Respir J, vol. 19, pp. 121- 6, 2002.

[94]. V.G. Kirk, S.G. Bohn, W.W. Flemons, and J.E. Remmers, “Comparison of home oximetry monitoring with laboratory polysomnography in children,” Chest, vol. 124, pp. 1702-8, 2003.

[95]. Y. Fischer, B. Junge-Hülsing, G. Rettinger, and A. Panis, “The use of an ambulatory, automatic sleep recording device (QUISI version 1.0) in the evaluation of primary snoring and obstructive sleep apnoea,” Clin Otolaryngol, vol. 29, no. 1, pp. 18 – 23, 2004.

[96]. K.P. Pang, and D.J. Terris, “Screening for obstructive sleep apnea: an evidence-based analysis,” Am J Otolaryngol, vol. 27, pp. 112-8, 2006.

[97]. J.H. Ficker, G.H. Wiest, Wilpert J, F.S. Fuchs, and E.G. Hahn, “Evaluation of a portable recording device (Somnocheck) for use in patients with suspected obstructive sleep apnoea,” Respiration, vol. 68, no. 3, pp. 307-12, 2001.

[98]. N.T. Ayas, S. Pittman, M. MacDonald, and D.P. White, “Assessment of a wrist-worn device in the detection of obstructive sleep apnea,” Sleep Med, vol. 4, pp. 435-42, 2003.

Appendix A

Description of features in the best feature combinations for OSA screening and OSA severity estimation

A.1 Best set of features for OSA screening

- Spectral centroid of PSD for the band of 450-600 Hz, for expiration phase of mouth breathing in upright posture
- Spectral bandwidth for the band of 1200-1800 Hz, for expiration phase of mouth breathing in supine posture
- Difference between nose and mouth breathing in spectral centroid for the band of 100-2500 Hz, for expiration phase in supine posture
- Difference between supine and upright postures in signal power for the band of 100-150 Hz, for inspiration phase of mouth breathing
- Bispectral invariant at 15°, for inspiration phase of mouth breathing in upright posture
- Bispectral invariant at 24°, for inspiration phase of mouth breathing in upright posture

- Bispectral invariant at 27°, for inspiration phase of mouth breathing in upright posture

A.2 Best set of features for OSA severity estimation

- Signal power in the band of 450-600 Hz, for inspiration phase of nose breathing in supine posture
- Signal power in the band of 600-1200 Hz, for expiration phase of mouth breathing in supine posture
- Crest factor for the band of 100-2500 Hz, for inspiration phase of nose breathing in upright posture
- Spectral centroid for the band of 600-1200 Hz, for inspiration phase of nose breathing in upright posture
- Spectral centroid for the band of 600-1200 Hz, for inspiration phase of mouth breathing in supine posture
- Spectral centroid for the band of 1200-1800 Hz, for inspiration phase of nose breathing in upright posture
- Spectral bandwidth for the band of 100-150 Hz, for inspiration phase of mouth breathing in supine posture
- Spectral bandwidth for the band of 150-450 Hz, for expiration phase of nose breathing in upright posture
- Spectral bandwidth for the band of 120-1800 Hz, for expiration phase of mouth breathing in supine posture
- Difference between nose and mouth breathing in signal power in the band of 600-1200 Hz, for expiration phase of upright posture

- Difference between nose and mouth breathing in relative signal power for the band of 1800-2500 Hz, for expiration phase in supine posture
- Difference between nose and mouth breathing in spectral centroid for the band of 150-450 Hz, for expiration phase in supine posture
- Difference between nose and mouth breathing in spectral bandwidth for the band of 100-2500 Hz, for inspiration phase in upright posture
- Difference between nose and mouth breathing in spectral bandwidth for the band of 1200-1800 Hz, for inspiration phase in upright posture
- Difference between supine and upright postures in signal power for the band of 100-150 Hz, for inspiration phase of nose breathing
- Difference between supine and upright postures in relative signal power for the band of 1800-2500 Hz, for inspiration phase of nose breathing
- Difference between supine and upright postures in median bi-frequency in f_2 direction, for inspiration phase of nose breathing
- Area under the bispectrum in [350-600 Hz, 2350-2600 Hz], for inspiration phase of nose breathing in supine posture
- Area under the bispectrum in [850-1100 Hz, 1600-1850 Hz], for inspiration phase of nose breathing in upright posture
- Area under the bispectrum in [350-600 Hz, 600-850 Hz], for inspiration phase of mouth breathing in upright posture
- Difference between nose and mouth breathing in are under the bispectrum in [850-1100 Hz, 1600-1850 Hz], for inspiration in upright posture
- Difference between supine and upright postures in are under the bispectrum in [1850-2100 Hz, 2100-2350 Hz], for inspiration phase in nose breathing