# Acquisition and analysis of snore signals for diagnosis of obstructive sleep apnea

Ng, Andrew Keong

2010

# ACQUISITION AND ANALYSIS OF SNORE SIGNALS

# FOR DIAGNOSIS OF OBSTRUCTIVE SLEEP APNEA

**NG KEONG ANDREW**

**School of Electrical & Electronic Engineering**

A thesis submitted to the Nanyang Technological University

in fulfillment of the requirement for the degree of

Doctor of Philosophy

**2010**

*To my family,*

*for their love and support.*

# Acknowledgements

Although there is only one name on the title page, the realization of this thesis would not have been possible without the guidance and support of many others.

First and foremost, I would like to express my deepest appreciation to Associate Professor Koh Tong San, my thesis advisor, for his consistent guidance and invaluable suggestions throughout the course of this research work. He has always been there to help me despite of his busy schedule.

I would also like to thank Dr. Kathiravelu Puvanendran, my clinical collaborator, for his unwavering support that has allowed me to conduct clinical studies involving human subjects in the sleep laboratory.

I am grateful to Dr. Eugene Baey, my industrial collaborator, for his constant encouragement and professional insights into advanced respiratory technologies and innovations.

I am also very thankful to the polysomnographic technicians in the Sleep Disorders Unit of Singapore General Hospital and the signal processing specialists in our institution for their active participation, as well as their innumerable clinical and technical assistance.

Last but not least, my greatest debt of gratitude goes to my family, especially my parents and my aunt, whose hard work and sacrifices gave me the opportunities I have today. For their unconditional love and support, and everything else, I dedicate this thesis to them.

# Table of Contents

# Abstract

Obstructive sleep apnea (OSA) is a sleep-related breathing disorder that is common worldwide and potentially life-threatening; however, many affected individuals remain undiagnosed and untreated. This research aims to innovate on a simple, cost-saving, and reliable approach to diagnose OSA via the acquisition and analysis of snore signals, with an intention to mass screen for OSA. This thesis attempts to achieve the research aim through: (1) the implementation of a robust and user-friendly acquisition system for snore signals, along with recommendations for measurement standards; (2) the development of an advanced wavelet-driven preprocessing system that efficiently integrates both snore signal enhancement and snore activity detection; (3) the identification of effective snore-based OSA diagnostic markers, including formant frequencies (82.5–100% sensitivity, 82.0–95.0% specificity), wavelet bicoherence peaks (82.5–100% sensitivity, 83.3–100% specificity), and psychoacoustic metrics (72.0–78.0% sensitivity, 91.2–92.0% specificity), which accurately classify apneic and benign snores in same- and both-gender patient groups (p-value < 0.0001); (4) the formulation of regression models that are indicative of OSA severity; (5) the investigation of physiological-anatomical-acoustical relationships of snores via source-filter modeling; and (6) the successful generation of natural-sounding synthetic snores using a novel snore source flow model. Results consistently reveal that snore signals carry rich information for OSA detection; therefore, the use of snore properties to distinguish between patients with and without OSA is promising. Continued exploration in this research area will certainly unfold the clinical value of snore signals in diagnosing OSA in the near future.

# Summary

Obstructive sleep apnea (OSA) is a common and potentially devastating sleep-related breathing disorder characterized by a cessation of respiratory for at least 10 s, corresponding to a complete upper airway (UA) occlusion despite continuous chest and abdominal movements. If left untreated, patients with OSA can suffer from many serious health issues including cardiovascular and cerebrovascular morbidity and mortality, excessive daytime somnolence and fatigue, and significant societal costs. The gold standard for diagnosing OSA is an overnight multichannel polysomnography, which is expensive, time-consuming, and labor-intensive to perform, leading to high undiagnosis and untreated rate. Hence, many researchers have attempted to explore other modalities (e.g., oxygen saturation regularity, nasal airflow pressure, heart rate variability, thoracic body movement, and snore acoustical properties) to detect OSA. Among these modalities, snore-based analysis has received growing interests as it can potentially provide simple, inexpensive, time- and labor-saving, safe, and rapid screening for OSA.

Snoring is a typical and earliest symptom of OSA, affecting 70–95% of patients with OSA worldwide. It is caused by the vibrations of soft tissues and/or turbulence of airflow at constrictions in the UA. Therefore, any structural changes in the UA, which acts as an acoustic filter, may reflect in the acoustical properties of snores, facilitating the discrimination between apneic and benign patients. Based on this underlying hypothesis, several snore-based diagnostic methodologies (e.g., snore counting, sound intensity measurement, power spectrum estimation, and pitch calculation) have been presented in the literature. While these methodologies lend support to the hypothesis

that snore-based analysis has the potential to diagnose OSA, they produce inconsistent, contradictory, and unconvincing outcomes, which can be attributed to inappropriate instrumentation and measurement, as well as ineffective analysis algorithms.

In this thesis, we proposed several robust and effective approaches on snore signal acquisition and analysis for the diagnosis of OSA. A high-fidelity and user-friendly acquisition system for snore signals is developed and implemented in a sleep laboratory (Sleep Disorders Unit, Singapore General Hospital, Singapore). System components (microphone, preamplifier, data acquisition card, grounding layout, and graphical user interface) were carefully chosen and installed; consequently, background acoustical noise and electromagnetic interference are substantially reduced, yielding snore signals of better quality and integrity. A mouth-to-microphone distance of 0.3 m can also achieve a good tradeoff between signal quality and patient comfort, bringing more standardization and consistency to the acquisition process.

To further improve signal quality and intelligibility, we devised a novel wavelet-driven preprocessing system that elegantly incorporates both snore enhancement and detection in a translation-invariant discrete wavelet transform domain. The proposed level-correlation-dependent (LCD) threshold and snore activity detector within the preprocessing system are highly comparable to the conventional denoising techniques via level-dependent threshold and the segmentation methods via short-time energy and zero-crossing rate, capable of achieving the best outcome in the objective and subjective experiments conducted, with a snore signal-to-noise ratio improvement of 3.02–38.22 dB and a snore activity detection accuracy of 50.63–97.47%. Statistical quality of the LCD threshold was also proven theoretically. Clinical efficacy of the preprocessing system was assessed, and results show that snores analyzed after preprocessing deliver higher OSA diagnostic accuracy than those before preprocessing.

Subsequently, the preprocessed snores from apneic and benign patients, correspondingly defined as apneic snores (AS) and benign snores (BS), were analyzed by means of linear prediction (LP), discrete all-pole modeling (DAP), wavelet bicoherence (WBC), and even psychoacoustic metrics. Useful acoustical characteristics of snore signals were extracted and identified as OSA diagnostic markers. These markers include the first formant (F1) in the LP or DAP spectrum, the peak frequency at $f_1$ (PF1) and the peak sum frequency (PSF) in their respective WBC spectrum, as well as the loudness and the psychoacoustic annoyance.

Diagnostic performance of F1, PF1, and PSF were appraised using 30 training and 10 test data from each of the 40 participating patients (30 apneic and 10 benign). Quantitative differences in F1, PF1, and PSF between AS and BS are found in all three patient groups (males, females, and both males and females combined), giving rise to high sensitivity (F1: 82.5–100%; PF1: 85.4–98.3%; PSF: 82.5–100%) and high specificity (F1: 82.0–95.0%; PF1: 85.0–100%; PSF: 83.3–100%) with statistical significance (p-value < 0.0001). These diagnostic results outperform those obtained from the widely used spectral peak frequency (PF) whose sensitivity and specificity are 62.5–91.7% and 70.0–97.5%, respectively. Relationship between apnea-hypopnea index (AHI, i.e., OSA severity measure) and the proposed diagnostic markers can possibly take the functional form of exponential or power, yielding a predictive AHI value (F1: 10.3–14.9 events/h; PF1: 8.0–9.8 events/h; PSF: 9.3–10.7 events/h) that is close to the ideal value of 10.0 events/h.

Apart from the above, loudness and annoyance are found to be the best two among five psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) in the classification of AS and BS. While both the metrics indicate statistically significant differences between the diagnostic classes (p-value <

0.0001), loudness achieves a higher sensitivity (78.0% versus 72.0%) but lower specificity (91.2% versus 92.0%) than annoyance. Nonetheless, both the metrics exhibit stronger correlation with AHI than with neck circumference (NC) or with body mass index (BMI); quantitatively, Pearson's product-moment and Spearman's rank correlation coefficients are respectively $r_p$ = 0.7182–0.7432 and $r_s$ = 0.7062–0.7162 for AHI; $r_p$ = 0.5313–0.5564 and $r_s$ = 0.6638–0.6828 for NC; and $r_p$ = 0.4356–0.4963 and $r_s$ = 0.5340–0.5878 for BMI, with p-values < 0.0001.

To validate the reliability of the OSA diagnostic markers and to understand the mechanisms of snore production, we investigated the relationships between the snore source flow (SF, i.e., acoustic source in snore production), the UA anatomical structures, and the snore characteristics by means of an area perturbation study and a SF analysis. The former study examines the acoustical and perceptual impacts of changing the cross-sectional areas (CSA) of the pharynx (PX) and oral cavity (OC) on the generation of snores, whereas the latter analysis parameterizes and models SF and its derivative (SFD), along with the generation of synthetic snores using various SFD pulse shapes.

Results reveal that alterations in the CSA of PX and OC have more implications on the acoustical than the perceptual aspect of snores. F1, but not PF, consistently increases with narrowing pharyngeal airway, being higher for a wider mouth opening. Accordingly, F1 is more capable of distinguishing apneic and benign patients than PF due to the close association of F1 with the UA anatomical structures, as evidenced in the earlier diagnostic results. In contrast, there are no substantial differences in snore sound perception after narrowing or widening the pharyngeal airway and mouth opening; however, these CSA alterations can indirectly affect the psychoacoustics of snore sounds by changing the SF waveforms. Adding weight to these findings,

changes in the shape of SF or SFD pulse can influence the acoustical and perceptual characteristics of snores. The SF pulse shapes are different among snores and can be related to the biomechanical properties (e.g., compliance and elasticity) of snore excitation source (ES), further suggesting that the temporal and spectral attributes of SF and SFD may possess important information about the dynamical behavior of ES. In addition, we proposed a SFD model derived from the second derivative of the Gaussian probability density function or the Mexican hat wavelet with effective compact support of [-5, 5] and fruitfully generated natural-sounding synthetic snores, thereby offering a greater understanding of physiological, anatomical, acoustical, and perceptual perspectives on snore production mechanisms.

In conclusion, this thesis demonstrates that snore properties contain essential information for detecting OSA; the use of snore signals as a simple, inexpensive, safe, and reliable approach to diagnose OSA is feasible. Given these encouraging results from snore signal acquisition and analysis, we strongly believe that continuation in this research area will unveil the clinical value of snore signals in the near future.

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| AHI | apnea-hypopnea index |
| AS | apneic snores |
| AUC | area under receiver operating characteristic curve |
| BMI | body mass index |
| BS | benign snores |
| CAS | cross-sectional areas |
| CWT | continuous wavelet transform |
| DAP | discrete all-pole |
| DWT | discrete wavelet transform |
| EMI | electromagnetic interference |
| ES | excitation source |
| FFT | fast Fourier transform |
| F1 | first formant |
| F2 | second formant |
| F3 | third formant |
| LCD | level-correlation-dependent |
| LD | level-dependent |
| LP | linear prediction |
| MMD | mouth-to-microphone distance |
| MOS | mean opinion score |
| NC | neck circumference |
| OC | oral cavity |

| | |
|---|---|
| OSA | obstructive sleep apnea |
| PF | peak frequency |
| PF1 | peak frequency at $f_1$ |
| PSD | power spectral density |
| PSF | peak sum frequency |
| PSG | polysomnography |
| PX | pharynx |
| QPC | quadratic phase coupling |
| ROC | receiver operating characteristic |
| RSD | residual standard deviation |
| SA | snore activity |
| SE | standard error of area under receiver operating characteristic curve |
| SF | source flow |
| SFD | source flow derivative |
| SNR | signal-to-noise ratio |
| TIDWT | translation-invariant discrete wavelet transform |
| UA | upper airway |
| WBC | wavelet bicoherence |

## Chapter 1

# Introduction

This thesis focuses on the acquisition and analysis of snore signals for the diagnosis of obstructive sleep apnea (OSA), a common sleep-disordered breathing. The present chapter provides the background of this research work, including the pathophysiology and complications of OSA, the pathophysiology of snoring, as well as the current gold standard diagnostic tool for OSA, namely polysomnography (PSG). Motivations and objectives of the work are subsequently identified, together with the rationale for using snore signals to detect OSA. We concluded this chapter by summarizing the novel contributions of this work to both the engineering and medical fields.

## 1.1    Background

### 1.1.1    Obstructive Sleep Apnea

OSA is the most prevalent condition among the sleep-related breathing disorders, affecting 9–24% of men and 4–9% of women in the United States population aged 30–60 years [1], with no major discrepancies found between African-Americans and White people [2]. In Singapore, 21.1% of the middle-aged population are suspected of harboring OSA [3]. Among adults with OSA, the occurrence of OSA with daytime hypersomnolence in the United States is 4–5% and 2–3% in men and women, respectively [1], which is comparable to the affliction in Asian men and women that ranges from 4.1% to 7.5% and 2.1% to 3.2% , respectively [4].

The pathophysiology of OSA relies on the balance of forces between the

_____

The contents of this chapter are partly based on the author's publication [A11].
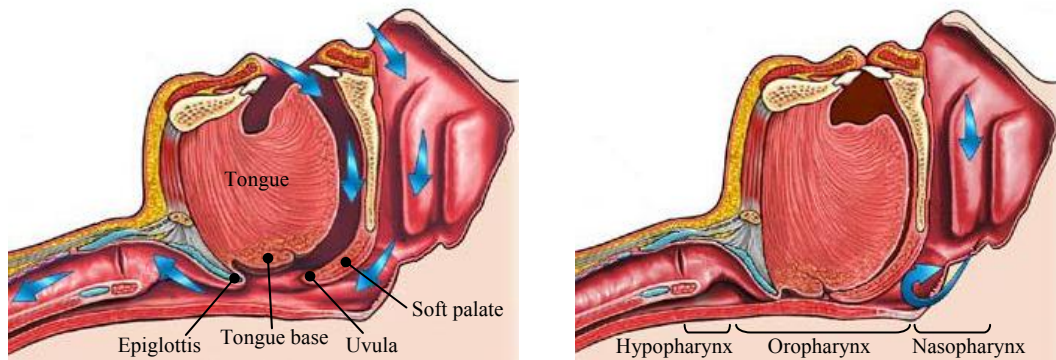
**Figure 1.1**   Pictorial illustrations of a normal airway (left) and an obstructed airway (right) [5].

upper airway (UA) dilator muscles (e.g., genioglossus and tensor palatini) that maintain airway patency and the negative pharyngeal intraluminal pressure created by thoracic expansion during inspiration [6,7]. When the tonic contraction and tension of the dilator muscles fail to counteract the negative intraluminal pressure, the soft pharyngeal walls collapse and occlude the UA, as illustrated in Figure 1.1. Explicitly, OSA is characterized by repetitive sleep-related posterior pharyngeal obstruction for at least 10 seconds (s) with attendant oxyhemoglobin desaturation of more than 3% and heart rate reduction in spite of continuous chest and abdominal movements [7,8]. Recurrent airway occlusion results in an increased corrective inspiratory effort, which often triggers a transient arousal with sonorous respirations.

Far more than most people realize, OSA is a serious and potentially life-threatening condition. Most patients with OSA have symptoms related to snoring, choking at night, and frequent arousals [8-10]. Attributed to the intermittent hypoxemia and arousal-associated sleep fragmentation, they also experience daytime hypersomnolence and fatigue, cognitive dysfunction, as well as psychological disturbances [11,12], increasing risks of industrial accidents and driving fatalities [13,14]. If left untreated, the pathophysiological consequence of OSA can lead to cardiovascular and cerebrovascular morbidity and mortality [15,16], such as

hypertension [17], coronary artery disease [18], congestive heart failure [19], and cerebral vascular accidents [20]. Because of these symptoms and functional impairments, OSA patients frequently complain having a poor quality of life in emotional, physical, and social domains [21,22].

### 1.1.2    Snoring

Snoring, at the mild end of sleep-disordered breathing spectrum, is widespread in the general population, with a prevalence of 35–45% in men and 15–28% in women [1,23]. The incidence of snoring in both genders increases with age, being higher in men than women. In a 30- to 35-year-old population, 20% of men and 5% of women snore; by the age of 60 years, 60% of men and 40% of women snore habitually [24]. More than just a nocturnal nuisance, snoring is the earliest and most consistent symptom of OSA, occurring in 70–95% of patients with OSA worldwide [10]. In Singapore, 87.5% of 106 adult habitual snorers studied have OSA [3], and correspondingly, 72% of them suffer from OSA with daytime hypersomnolence.

Snoring is caused by the vibrations of soft tissues (e.g., soft palate, uvula, tonsils, tongue base, epiglottis, and lateral pharyngeal walls) and/or turbulence of airflow at constrictions in the UA [25]. According to the Bernoulli's principle [26,27], when a constant airflow passes through the narrow pharynx, the airflow velocity increases because of mass conservation, while the intraluminal pressure decreases. This phenomenon further diminishes the pharyngeal airway size and promotes its occlusion. Any imbalance of forces between the UA dilator muscles and the negative intraluminal pressure will cause soft tissues in the UA to vibrate, which indicates a partial airway obstruction, and/or trigger an OSA attack. Figure 1.2 highlights the resemblance between the pharyngeal shape and the Venturi tube, where the effect of
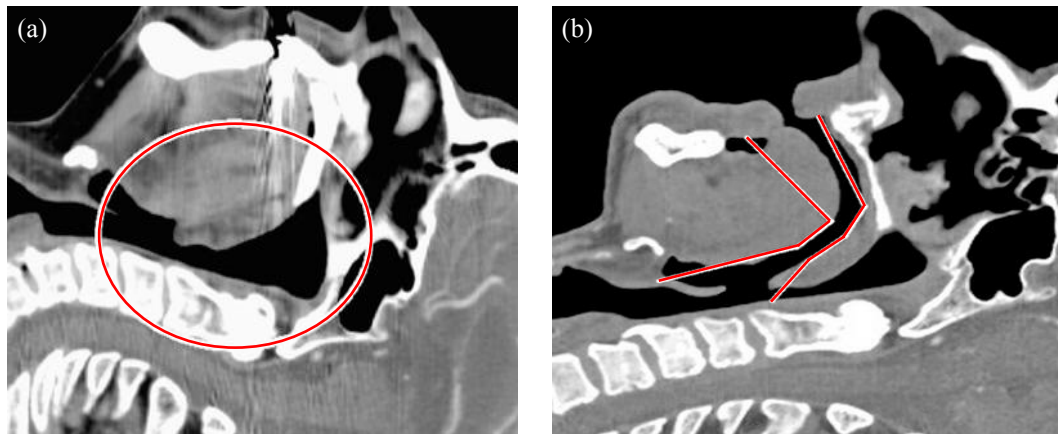
**Figure 1.2**   Computed tomography images of upper airway highlighting Venturi tube shape of the pharynx for a (a) nose breather and (b) mouth breather [27].

Bernoulli's principle can take place, for both nose and mouth breathers.

A mathematical model of snoring was developed by means of airflow in an elastic tube [28]. When the geometry, resistance, compliance, elasticity, and airflow of the model are appropriately varied, the tube opens and closes repetitively, producing sound similar to snoring. When these variables are further altered, the tube completely collapses and blocks the flow of air, simulating OSA. This model further suggests that snoring is a hallmark of OSA [8,25], and anatomical and functional abnormalities in the UA (e.g., elongated soft palate, wide tonsillar pillars, redundant pharyngeal mucosa, increased pharyngeal airway narrowing, inadequate pharyngeal dilator muscle tone, and inflammation of pharyngeal walls) play important roles in the development and severity of OSA [29-31].

### 1.1.3    Polysomnography

Difficulties in the diagnosis of OSA based on patient's clinical history and physical examination have led to the landmark development of PSG in 1974 to describe the recording and interpretation of multiple simultaneous physiological parameters
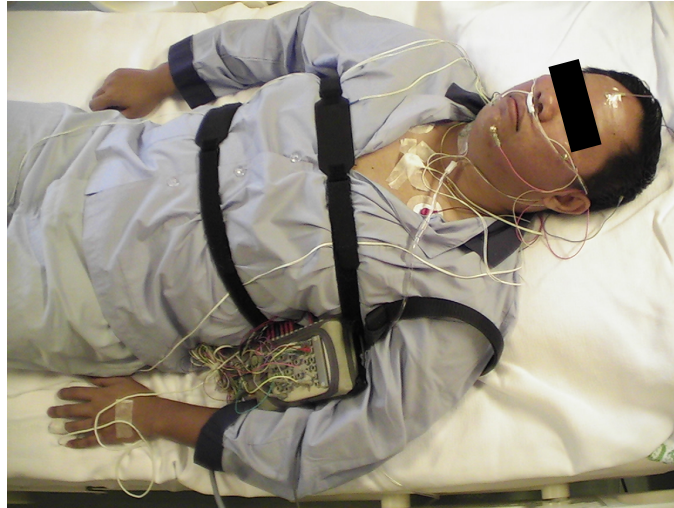
**Figure 1.3**   Patient prepared for nocturnal polysomnography.

associated with sleep [32]. Being a gold standard diagnostic tool, PSG provides a detailed assessment of biophysiological changes that happen during sleep. Consequently, a polysomnographic recording normally lasts for 6–8 hours (h) of regular nocturnal sleep and incorporates various electrophysiological and cardiorespiratory parameters, for instance, electroencephalogram, electroocculogram, electromyogram, electrocardiogram, thoracic and abdominal respiratory effort, nasal and oral airflow, snoring level, pulse rate, oxygen saturation, body position, as well as limb movement [33], as depicted in Figure 1.3.

Interpretation of PSG is a tedious process as it engages in both qualitative and quantitative measures of sleep quality and respiratory effort. These measures include, but are not limited to, sleep onset latency, sleep stage time, sleep efficiency, number of arousals per hour, number of respiratory disturbances per hour, and rate of oxyhemoglobin desaturation. Among them, apnea-hypopnea index (AHI) is widely used as an indicator of OSA and a proxy measure of OSA severity; it is expressed as total number of apneas and hypopneas per hour of sleep (events/h), in which a hypopnea is defined as a 30–50% reduction in respiratory airflow for at least 10 s with

greater than 3% oxyhemoglobin desaturation [7,8,34]. Relative to a hypopneic event, an apneic event is more intense and dangerous because it involves complete stoppage of airflow into and out of the lungs. Clinically, a subject can be considered as suffering from OSA if AHI ≥ 10 events/h, which can be further classified into mild (AHI = 10–15 events/h), moderate (AHI = 15–30 events/h), and severe (AHI > 30 events/h) [34].

A sample polysomnographic report of an apneic (AHI ≥ 10 events/h) and a benign (AHI < 10 events/h) patient who underwent nocturnal PSG at the Sleep Disorders Unit of Singapore General Hospital in Singapore, where we were collaborating with for this research, are included in Appendices A and B, respectively.

## 1.2    Motivations

Despite the detrimental effects of OSA, ranging from physical and emotional health to social functioning, nearly 82% of affected men and 93% of affected women in the United States remain undiagnosed and untreated [35,36], which augments the burden of this disorder. Case reports have asserted a high incidence of postoperative complications [37,38] and deaths [39,40] among patients with undiagnosed OSA. In addition, the severity of OSA is associated with the magnitude of medical costs; OSA confers an approximate two-fold increase in medical expenses in the years preceding the diagnosis, and treating the condition lowers the healthcare costs [41,42]. Therefore, untreated OSA can be both deleterious to the patients and expensive to the healthcare economics, amounting to billions of dollars per year [43]. The threats of OSA to patient health and global economy have fully justified the need for early recognition of subjects at risk of OSA. If patients are identified and treated at an initial stage of OSA, the adverse health effects will be substantially minimized and postoperative complications will be prevented [44,45].

A dominating reason for the sky-scraping undiagnosis rate is the overnight multichannel PSG, which is not only time-consuming and inconvenient, but also labor-intensive and costly. Patients are required to stay overnight in a sleep laboratory, away from family, tethered to a multitude of physiological instruments measuring electrophysiological and cardiorespiratory parameters, as pictured in Figure 1.3. Owing to its complex setup, recording, and scoring protocols, PSG can only be performed by trained polysomnographic technicians [33], which limits its applications. Besides that, the cost of a polysomnographic test, spanning from US$800 to US$1400, is hardly affordable for the general public [46,47]. In Singapore, it can cost between S$511 and S$1335, or an equivalent of US$340–US$887.

The shortage of equipped sleep laboratories and qualified sleep physicians is another drawback of PSG and has resulted in lengthy waiting lists. The wait time can be as long as 10 weeks [48] or even longer in some areas (e.g., Veterans Affairs Medical Centers) [49,50], causing delays that heighten the risk of accidental injury or death to oneself and others [51-53]. Long waiting times also contribute to elevated no-show rates that indirectly extend the waiting lists [54]. A study has noted that at least 2310 PSGs per 100000 population per year are required to adequately alleviate the problem of undiagnosed OSA; this exceeds by a factor of 10 in most countries and a factor of 50 in the United Kingdom [50]. Apparently, PSG is not readily available to accommodate the growing demand for OSA diagnosis.

It is these drawbacks of PSG that motivate us to explore other modalities that can potentially offer convenient, inexpensive, and rapid screening for OSA. Snoring sounds are readily available in most subjects with suspected OSA, and digital analysis of snore signals provides numerous benefits. First, snoring sounds can be easily recorded at the patient's home or in a sleep laboratory via a portable audio tape

recorder, thereby providing cost-saving and non-invasive way of instrumentation and data measurement. No technical expertise is also required during system setup and recording. Second, the acquired snore signals can be either sent back to the laboratory or uploaded to an online system for processing, leading to time- and labor-saving. As signal analysis and classification can be automatically carried out by computer software packages (e.g., LabVIEW$^{TM}$ and MATLAB$^{TM}$), errors and inconsistencies caused by the technician's subjectivity and fatigue during PSG scoring [55] will be minimized. Last but not least, with the aforementioned benefits of snore-based OSA diagnosis, large-scale and community-oriented screening for OSA is achievable, shortening the delays and allowing earlier treatment and better control of the condition.

We have sufficient clinical and technical evidences to support our hypothesis that acoustical properties of snore signals contain essential information to discriminate between patients with and without OSA. As mentioned in Section 1.1.2, fluttering vibrations of soft tissues and/or noise-like turbulent airflow at constructions in the UA produce acoustic waves [25]. While these waves propagate through the UA, they are spectrally modified by the UA anatomical structures (e.g., airway cross-sectional areas and airway length) to create distinct sounds of snoring before reaching the listener. It is well-recognized that patients with OSA usually possess smaller and more collapsible UAs than those without OSA because of enlarged soft tissues and decreased muscle tone [29-31,56,57]. The presence of abnormalities in the UA may alter the acoustical characteristics of simple or benign snores, just as the existence of laryngeal pathology changes normal vibratory patterns of the vocal folds and influences the quality of speech signals [58-67], thereby yielding distinctive acoustical signatures in snore signals that helpfully facilitate the classification between apneic and benign patients.
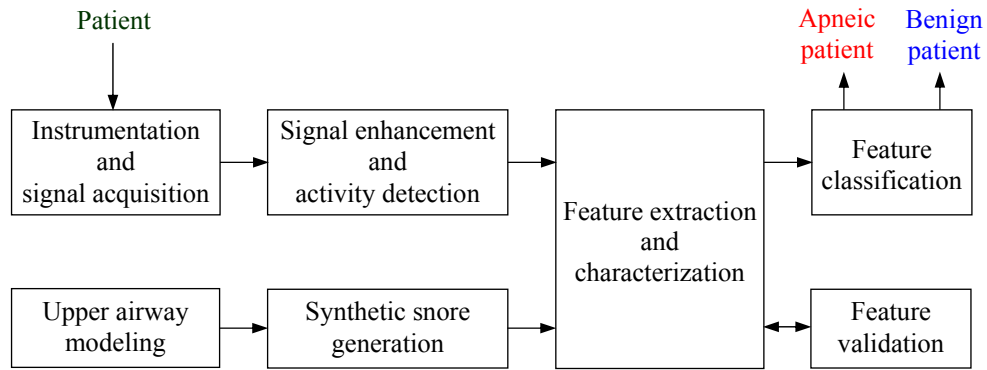
**Figure 1.4**   Research workflow.

## 1.3    Objectives

The overall aim of this research is to innovate on a simple, low-cost, safe, and reliable approach to diagnose OSA through the acquisition and analysis of snore signals, with an intention to mass screen for OSA. In order to achieve the aim, several objectives have to be met:

(a)    to develop and set up an acquisition system for snore signals;

(b)    to improve snore signal quality and detect snore activity;

(c)    to identify useful acoustical characteristics of snore signals as OSA diagnostic markers;

(d)    to classify snores produced by apneic and benign patients using the diagnostic markers;

(e)    to ascertain the correlation between AHI and the diagnostic markers; and

(f)    to investigate physiological-anatomical-acoustical relationships of snores for validation of the diagnostic markers and for understanding the mechanisms of snore production.

The block diagram in Figure 1.4 displays an overview of the workflow of this research process.

## 1.4    Major Contributions of the Thesis

In this research on snore-based diagnosis for OSA, we have made the following contributions:

(a)    We designed and implemented a high-fidelity and user-friendly acquisition system for snore signals in a sleep laboratory (Sleep Disorders Unit, Singapore General Hospital, Singapore) [A13].

(b)    We proposed standardization measures in the acquisition of snore signals, in particular the optimal mouth-to-microphone distance [A6].

(c)    We developed an efficient preprocessing system that simultaneously enhances snore signal-to-noise ratio in the range of 3.02–38.22 dB and detects snore activity with an accuracy of 50.63–97.47% within a translation-invariant discrete wavelet transform domain [A4,A10,A12].

(d)    We proposed several effective and reliable snore-based OSA diagnostic markers including formant frequencies [A2,A5,A14,A15,A16], wavelet bicoherence peaks [A3,A8], and psychoacoustic metrics [A2,A7,A9]. These diagnostic markers indicate statistically significant differences (p-value $< 0.0001$) between apneic snores (AS, i.e., snores produced by apneic patients) and benign snores (BS, i.e., snores produced by benign patients) in same-gender (i.e., males and females separately) and both-gender (i.e., males and females combined) patient groups, with sensitivity (i.e., the ability to correctly identify AS) and specificity (i.e., the ability to correctly identify BS) between 82.0% and 100% for the objective markers, and between 72.0% and 92.0% for the subjective markers.

(e)    We formulated regression models that can competently predict AHI cutoff value within 8.0–14.9 events/h, which is close to the ideal value of 10.0 events/h [A2,A3,A5,A8,A14,A15,A16].

(f)    We parameterized (both time- and amplitude-based parameters) and devised snore source flow model that successfully generates natural-sounding synthetic snores [A1].

## 1.5    Organization of the Thesis

This thesis is organized as follows:

Chapter 1 – Introduction [A11]

This chapter provides the background of the present research on snore-based OSA diagnosis, including the pathophysiology and complications of OSA, the pathophysiology of snoring, and the applications and drawbacks of PSG. Motivations and objectives of this work are identified, along with justifications for the research hypotheses. Novel contributions of this work are also highlighted.

Chapter 2 – Review of Snore-Based Diagnostic Methodologies [A11]

This chapter presents an in-depth literature review on the methodologies adopted for analyzing snore signals to distinguish between apneic and benign patients. The review compares and contrasts four main methodology categories, describing the methodologies and evaluating their advantages and disadvantages. Furthermore, gaps and omissions in the literature are outlined, and the emerging issues on the need for effective approaches to acquisition and analysis of snore signals for diagnosing OSA are addressed.

Chapter 3 – Snore Signal Acquisition System [A6,A13]

This chapter discusses the design and implementation of an acquisition system for snore signals in a sleep laboratory. Various aspects on the selection and installation of the system components (microphone, preamplifier, data acquisition card, grounding layout, and graphical user interface) are described, with experiments to appraise the robustness of the proposed acquisition system. The influence of microphone positioning on signal quality and patient comfort are also reported.

Chapter 4 – Snore Signal Preprocessing System [A4,A10,A12]

This chapter proposes an advanced preprocessing system that concurrently enhances signal quality and detects snore activity via translation-invariant discrete wavelet transform. Detailed formulations of the system algorithms (level-correlation-dependent threshold and snore activity detector) are shown, and the comparison methods and evaluation metrics are outlined. Experimental results, in terms of theoretical statistics quality, signal-to-noise ratio, mean opinion score, and clinical efficacy, are also presented and discussed.

Chapter 5 – Parametric Analysis and Classification of Snores [A2,A5,A14,A15,A16]

This chapter introduces a parametric approach to analyze snore signals and classify AS and BS in same- and both-gender patient groups. Formant frequencies, extracted from both linear prediction and discrete all-pole spectra, are proposed. Diagnostic accuracy of this approach is discussed using receiver operating characteristic curves and notched box plots. Moreover, linear and nonlinear regression models illustrating the relationship between AHI and the proposed markers are identified.

Chapter 6 – Nonlinear Analysis and Classification of Snores [A3,A8]

This chapter examines the feasibility of using nonlinear coupling between frequency modes in snore signals for classifying AS and BS in same- and both-gender patient groups. Two diagnostic markers derived from frequency modes with high nonlinear coupling strength in their respective wavelet bicoherence spectrum are proposed. Diagnostic accuracy of this approach is discussed, together with regression models explaining the relationship between AHI and the proposed markers.

Chapter 7 – Psychoacoustic Analysis and Classification of Snores [A2,A7,A9]

This chapter explores the application of psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) to snore sounds for the classification of AS and BS. Experimental conditions, including the listening test procedure and the rating scale, are described. Results are quantitatively presented using receiver operating characteristic curves and notched box plots, as well as Pearson's product-moment and Spearman's rank correlation coefficients to valuate the association between the psychoacoustic metrics, body mass index, neck circumference, and AHI.

Chapter 8 – Snore Physiological-Anatomical-Acoustical Relationships [A1,A2]

This chapter investigates the acoustical and perceptual impacts of changing the cross-sectional areas of the pharynx and oral cavity on the production of snores. Besides that, it parameterizes and models snore source flow, with application to synthetic snore generation. Both objective and subjective experimental results for the area perturbation study and the snore source flow analysis are presented and discussed with graphical illustrations.

**Table 1.1**   Demographic and clinical data of patients participated in research.

| Type | Sample size | Age (years) | BMI (kg/m$^2$) | NC (cm) | AHI (events/h) |
|---|---|---|---|---|---|
| Males    : Apneic | 24 | 41 ± 12 (23–77) | 28.9 ± 5.9 (18.2–42.8) | 41.3 ± 3.7 (32.0–51.0) | 52.8 ± 24.7 (11.6–101.9) |
| Males    : Benign | 6 | 36 ± 11 (20–52) | 28.1 ± 6.2 (21.6–37.4) | 40.3 ± 2.7 (37.0–44.0) | 5.3 ± 3.5 (0.2–8.9) |
| Females  : Apneic | 6 | 54 ± 14 (32–75) | 30.7 ± 10.5 (21.5–50.4) | 36.7 ± 4.7 (33.0–46.0) | 23.6 ± 15.1 (12.0–50.8) |
| Females  : Benign | 4 | 50 ± 6 (45–59) | 25.3 ± 4.7 (22.2–32.3) | 36.8 ± 2.8 (34.0–40.0) | 3.4 ± 3.2 (0.6–7.9) |
| Combined : Apneic | 30 | 44 ± 13 (23–77) | 29.3 ± 6.9 (18.2–50.4) | 40.4 ± 4.3 (32.0–51.0) | 46.9 ± 25.7 (11.6–101.9) |
| Combined : Benign | 10 | 41 ± 12 (20–59) | 26.9 ± 5.6 (21.6–37.4) | 38.9 ± 3.1 (34.0–44.0) | 4.6 ± 3.4 (0.2–8.9) |

Values are presented as mean ± standard deviation and range within brackets. BMI refers to body mass index in kg/m$^2$; NC, neck circumference in centimeters; AHI, apnea-hypopnea index in events/h; Combined, both males and females combined.

Chapter 9 – Conclusions and Recommendations

This chapter revisits the overall aim and objectives of this work. The research findings are summarized and related to the specific objectives. The limitations of this work are also highlighted with recommendations for future research.

The patient studies performed in this thesis were approved by the local Institutional Review Board. Written informed consent was obtained from all 40 participating patients, with their demographic and clinical data listed in Table 1.1, between January 2006 and January 2007. PSG recordings were manually scored following the internationally accepted Rechtschaffen and Kales criteria [68]. All analysis in this research work was performed within MATLAB$^{TM}$ (version 2006a, MathWorks Incorporated), unless otherwise stated.

## Chapter 2

# Review of Snore-Based Diagnostic Methodologies

As discussed in Section 1.2, OSA negatively influences one's quality of life in both health and socioeconomic factors, yet more than 80% of the affected individuals are undiagnosed and untreated [35,36]. This high undiagnosis rate can be attributed to the drawbacks of PSG (e.g., time-consuming, cumbersome, laborious, expensive, long waiting times, and limited availability). As a result, many researchers have attempted to search for other modalities to detect OSA. The modalities include oxygen saturation regularity [69-72], nasal airflow pressure [73-76], heart rate variability [77,78], and thoracic body movement [79,80]. Although these modalities can achieve a significant diagnostic accuracy of more than 80%, they require physical contact sensors, which may cause discomfort to the sleeping patients, especially for the pediatric patients. In addition, technical expertise is needed at the test site for correct placement of sensors and to monitor physiological data. Snore signals, on the other hand, can be non-invasively captured, and processing of snore signals can offer numerous benefits (e.g., convenient, inexpensive, time- and labor-saving, as well as readily suitable for mass screening).

Allured by the benefits of processing snores, various methodologies adopting acoustical properties of snore signals to distinguish between apneic and benign patients have been proposed over the past decade. These methodologies can be

_____

The contents of this chapter are partly based on the author's publication [A11].

grouped into four main categories: snore counting, sound intensity measurement, power spectrum estimation, and pitch calculation. This chapter presents a critical review of the literature. Throughout the review, we scrutinized each of the categories, describing the methodologies and evaluating their advantages and disadvantages. Furthermore, we pointed out gaps and omissions in the literature, and subsequently, we underscored the need for effective approaches to acquisition and analysis of snore signals for the diagnosis of OSA.

## 2.1    Snore Counting

The simplest methodology to quantify snoring is by recognizing its presence and counting the number of snores over a given time period. Snoring index, defined as average number of snoring events per hour, is implemented in a portable sleep-monitoring gadget for diagnosing OSA [81-83]. With the aid of the gadget, Brietzke *et al.* [81] examined 456 subjects with suspected OSA. Snoring sounds were captured by an electret microphone, which was attached to a cannula placed on the subject's upper lip with extensions to the nasal and oral apertures. A snoring event was characterized as any breath sound with non-uniform spectral pattern. Results demonstrated that the snoring index was poorly correlated with AHI (Spearman's rank correlation coefficient = 0.21, p-value < 0.0001). As mentioned in Section 1.1.1, snoring is interrupted by respiratory arrests when the UA dilator muscles fail to counteract the negative intraluminal pressure. Reestablishment of the UA patency often requires a transient arousal, accompanied by rapid sonorous respirations as the sufferer gasps for air [8-10]. In such a circumstance, the snoring index may only provide quantitative description of snoring frequency during sleep, but not fully reflect the pre- and post-apneic phases. Most importantly, it has no relation to the underlying pathophysiology

of OSA, thus lacking discriminatory power to determine the severity of OSA.

Apart from the snoring index, intermittent snoring is another predictive indicator of OSA. Penzel *et al.* [84,85] developed an ambulatory recording device that picks up snoring sounds using a laryngeal microphone. They interpreted a snore as any sound whose power within 50–800 Hz exceeded 50% of the total sound power, an intermittent snoring as any time interval of 5–60 s between two detected snores, and an intermittent snoring index (ISI) as number of intermittent snoring. Analysis of the ISI was performed on 68 patient datasets, and correlation between the ISI and AHI was found to be moderate (Pearson's product-moment correlation coefficient = 0.51). Reliability of the ISI was also validated by different researchers [86-88], and they noticed that the ISI could deliver high sensitivity (92–96%) but low specificity (16–27%). The statistical findings apparently imply that while most apneic patients are correctly diagnosed, many benign patients are falsely diagnosed as having OSA, causing them to receive unnecessary treatment and lengthening the delay into treatment for those who are in need.

## 2.2    Sound Intensity Measurement

Unlike counting number of snores, the methodology on sound intensity measurement stems from the fact that AS are louder and hoarser than BS. Wilson *et al.* [89] investigated the snoring sound intensity generated by 1139 subjects with sleep symptoms. Snoring sounds were recorded via a microphone suspended 0.6 metres (m) above the bed and subjectively assessed by technical personnel. Results highlighted that apneic patients produced significantly higher snoring sound intensity (> 5 dBA, A-weighted decibels) than benign ones (p-value < 0.0001), with a cutoff value of 38 dBA for an average sound intensity level to discriminate between the two diagnostic

classes. Moreover, the snoring sound intensity levels were associated with AHI, even after controlling for demographic and clinical factors. The study results are in agreement with those reported on pharyngeal function and snoring characteristics in apneic and benign patients [25,30]. Nonetheless, it is noteworthy to comment that sound intensity depends upon the distance between the subject's mouth and the microphone; if the distance is doubled, sound intensity will reduce by a factor of 4, or correspondingly, a 6 dB decrement in sound intensity level, according to the inverse square law [90]. Besides that, sound intensity measurement is susceptible to unwanted background acoustical noise, in particular, when the study provides no means of canceling the noise prior to further analysis and interpretation.

Utilizing both the snoring sound intensity and time interval for OSA screening, Brunt *et al.* [91] introduced an acoustical signature event (ASE) that was termed as any signal with peak above 100 µV, preceded by 10–90 s of silence. Snoring sounds were detected by a hanging microphone centered about 0.6 m above the head of the bed. Snore signals were amplified and channeled to an integration/rectification device before plotting them on a polygraph. The ASEs were manually scored, and an ASE index was formulated as total number of ASEs per hour of sleep. A total of 69 patients participated in the study; 63.8% of them were correctly classified into their diagnostic classes, with a sensitivity of 100% and a specificity of 19.3%, erring on the side of excessive false positives. The ASE index supports the anecdotal reports of AS and might potentially be an inexpensive screening tool for OSA. Conversely, determining the ASE index necessitates an overnight in-laboratory study and manual scoring; therefore, it is time-consuming and labor-intensive.

To improve the diagnostic accuracy of OSA, Issa *et al.* [92] additionally incorporated arterial oxygen saturation ($SaO_2$) data into the analysis of snore sound

intensity. They designed a digital recorder that collects snoring sounds from a transducer fastened to the patient's chest and $SaO_2$ from an ear oximeter. The collected data were transferred to a computer workstation and automatically analyzed. A snore was defined as any sound exceeding 0.8 V for more than 0.26 s, and an OSA attack was considered when two snores were separated by a silence of 10–120 s, along with a 3% decline in $SaO_2$. Based on 129 patient datasets, the recorder could diagnose OSA with a sensitivity of 84–90% and a specificity of 95–98%, exhibiting high predictive power in OSA diagnosis. Errors arising from subjective bias are also avoided since the analysis was automatically executed. In contrast, the chest transducer is sensitive to noise originated from chest movements, heart beats, and coughs. Furthermore, the additional use of oximetry increases both labor and medical costs.

## 2.3    Power Spectrum Estimation

A more technical methodology for detecting OSA is the use of spectral signatures in snore signals; the signatures are peak frequency (PF, i.e., frequency of highest peak in power spectrum), as well as frequency distribution of signal power. Hara *et al.* [93] employed a commercially available multidimensional voice program to examine acoustical characteristics of snoring sounds acquired from 58 subjects with symptoms suggestive of OSA. Sound recordings were done by a portable minidisc recorder through an omni-directional electret condenser microphone situated at 0.2 m above the subject's mouth. A snore was interpreted as any inspiratory breath sound. For each subject, 30 snores were windowed using the Hanning window function and transformed into frequency domain by the fast Fourier transform (FFT). Results indicated that AS have more high-frequency peaks than BS, with a mean PF of 408 Hz and 160 Hz for AS and BS, respectively, reaching a statistical significance of 0.012

(Mann-Whitney test).

Correlation between the PF and AHI was further explored by Michael *el al.* [94]. In their study, 60 patients were evaluated. An air-coupled microphone was positioned 0.3 m above the patient's head to pick up snoring sounds. Audio signals were processed via the FFT with a Hanning window of 16384 samples. The PF was measured for 3–5 inspiratory snores, and the mean value was computed for classification. Results showed that the PF significantly correlated with AHI (Spearman's rank correlation coefficient = 0.63, p-value < 0.0010). Spectral analysis of AS also revealed a non-harmonic noise pattern with high-frequency components above 1000 Hz, whereas BS were low-frequency patterns with harmonics occupying 100–300 Hz.

The abovementioned results for spectral analysis of AS and BS are consistent with the study outcomes of Perez-Padilla *et al.* [95]. In their study, snoring sounds were captured from two microphones, one attached to the patient's manubrium sterni, while the other was hung 0.15 m above the patient's head. A snore was described as any noise with audible low-frequency vibratory components. A sample of 19 patients was studied, and 10 snores from each patient were analyzed through the FFT. Results outlined that the PF is higher in AS than BS, with a cutoff value of PF = 500 Hz. To enhance the discrimination between AS and BS, they superimposed the power spectra of the 10 snores and inspected the frequency distribution of signal power. They observed that AS had relatively more power at high frequencies, and the ratio of power above 800 Hz to power below 800 Hz could be a predictive indicator of OSA. Given a power ratio of 0.3, it achieved a sensitivity of 40–88% and a specificity of 100%, depending on the snore selection; first post-apneic snore delivered better sensitivity than pre-apneic snores.

Despite the evidence that apneic patients generate snores with a higher frequency band than benign patients, the reliability of PF and power ratio in diagnosing OSA seems questionable because the work of Fiz *et al.* [96] yielded contradictory findings. Fiz *et al.* [96] examined the spectral signatures in snore signals of 17 subjects with suspected OSA. Snoring sounds were recorded using a miniature microphone placed on the subject's neck, 1 cm lateral to the midline at the level of the cricoid cartilage. Snores were recognized by listening, and power spectra of three consecutive snores from each subject were computed by means of the Welch's averaged modified periodogram algorithm with a 1024-sample Hanning window and a 50% overlap. Results illustrated that AS had a low PF with energy concentrated on a narrow frequency band (PF = 157 ± 136 Hz), while BS had a high PF with a fundamental frequency and harmonics (PF = 264 ± 107 Hz). Besides that, a negative correlation was noted between the PF and AHI (Spearman's rank correlation coefficient = -0.70, $p < 0.0016$), with a cutoff value of PF = 150 Hz to differentiate AS from BS.

The methodology on power spectrum estimation favorably reduces computational cost as only a certain number of snores are required for analysis and classification. Nevertheless, the diagnostic accuracy relies on the selection of snores, and there appears to be little consensus on the findings, weakening the impact of the methodology for the diagnosis of OSA.

## 2.4    Pitch Calculation

Instead of looking at the whole power spectrum to determine the PF or power ratio, recent researchers focused on the fundamental frequency or pitch of snore signals to screen for OSA. Abeyratne *et al.* [97] suggested an intra-snore-pitch-jump (ISPJ)

probability, which was expressed as the percentage of snoring episodes with pitch discontinuities, to distinguish AS from BS. A pair of hypercardioid microphone was sited at roughly 0.5 m above the patient's mouth to capture snoring sounds from the front, while rejecting sounds from 120º to the rear. Audio signals were high-pass filtered, and subsequently, snoring episodes and silent segments were estimated. A snoring episode was interpreted as a breath epoch containing a voiced-snoring segment, an unvoiced-snoring segment, and a silent segment, analogous to voiced and unvoiced speech [98]. The voiced- and unvoiced-snoring segments were processed through a pitch detector based on the center-clipping autocorrelation algorithm, and then the ISPJ probability were calculated from the voiced-snoring segments. A total of 45 patients took part in the study, and 25 snoring episodes from each patient were investigated. Depending on the pitch threshold setting and the number of pitch discontinuities exceeding a preset pitch threshold, the ISPJ probability could attain a sensitivity of 86–100% and a specificity of 50–80%. The ISPJ probability of snore signals might potentially be a predictor for OSA, if and only if, snore signals are inherently periodic, thereby limiting its applications.

Snores, especially AS, are generally produced by quasi- or non-periodic vibrations of soft tissues, together with a certain degree of noise created by turbulent airflow at constrictions in the UA [25,99,100]. Hence, snores with clear harmonics in the frequency domain are not commonly seen, as justified in the literature on power spectrum estimation of AS and BS [93-96]. The presence of irregularities in snore periodicity, as well as the incompetence of pitch-derived parameters for OSA screening are further evidenced in the work of Sola-Soler *et al.* [101].

Sola-Soler *et al.* [101] derived three parameters to quantify the pitch of snore signals, including mean value, standard deviation, and density of pitch, which the

latter was characterized as the fraction of snoring time where pitch was perceived over total snore duration. Snoring sounds were collected from 16 subjects at risk of OSA, with a piezoelectric contact sensor taped on the subject's neck beside the crycothyroid notch. The collected signals were amplified and band-pass filtered. Snores were detected by an automatic sound detector and subjectively assessed by a physician. Pitch detection was conducted on 683 snores via the center-clipping autocorrelation algorithm. Results highlighted that AS had higher pitch mean (19.55 Hz versus 14.30 Hz) and standard deviation (4.96 Hz versus 4.89 Hz) but lower pitch density (36.44 versus 39.21) than BS; however, these differences were not significant enough to separate between AS and BS by using one pitch-derived parameter due to large variability in snore frequency content. A linear discriminant analysis was also carried out with the parameters; it obtained a sensitivity of 58–59% and a specificity of 58–64%, depending on the number of independent variables. A monotonic decrease in the pitch density of AS with increasing AHI was found, which implies a greater irregularity in periodicity of AS. The statistical findings affirm that the pitch-derived parameters provide little usefulness towards the screening for OSA.

## 2.5    Discussion

The present chapter reviews four main categories of methodologies for identifying patients with and without OSA by means of snore signals. These methodology categories are: snore counting, sound intensity measurement, power spectrum estimation, and pitch calculation. Among them, the snore counting and the sound intensity measurement are the most straightforward methodologies since they can be easily implemented based on the temporal features of snore signals. No advanced signal processing algorithms is needed to analyze snores. The basic premise of these

two methodologies relies heavily on the definition of 'snore' and the acquisition of snore signals. To our best knowledge, there is neither an objective definition of snoring nor a standard protocol for measuring snoring sounds [25]. Consequently, as can be seen in the literature, different researchers adopted different interpretations for 'snore' and acquisition techniques for snore signals. The cutoff values and indices driven by these temporal features of snore signals are therefore restrained to the instrumentation and data acquisition techniques, lacking the robustness and the attributes for OSA detection. Although the incorporation of an arterial oxygen saturation data into snore sound intensity could boost the OSA diagnostic accuracy, the additional costs of having extra physiological sensors and technical expertise may defeat the purpose of exploring cheaper alternatives to PSG.

Unlike snore counting and sound intensity measurement, the methodologies on power spectrum estimation and pitch calculation apply the frequency information of snore signals in the detection of OSA. They involve only a handful of snores for analysis and classification, encouraging time-saving and computational efficiency. Moreover, no technical expertise is demanded at test site because the sound recording can be simply done through a non-contact microphone. Thus, these frequency-driven methodologies also promote labor-saving, community-oriented screening for OSA, and even home-based sleep studies. In spite of their advantages, the methodologies yield mixed, conflicting, and inconclusive results, diminishing its reliability and accuracy for diagnosing OSA.

There are two possible reasons leading to diagnostic failure for the frequency-driven methodologies. First, the FFT for power spectrum estimation may not completely reflect the UA acoustical features that capture the underlying pathophysiology of OSA, which is of supreme importance in the development of an

effective diagnostic predictor for OSA. Second, the autocorrelation algorithm with center clipping for pitch calculation is tailored for periodic signals, which may not be applicable to snore signals that are mostly quasi- or non-periodic [25,99,100].

The review of the literature accentuates the need for effective approaches, ranging from instrumentation and signal acquisition to snore analysis and classification, for the diagnosis of OSA. We attempted to fill the gaps and omissions in the literature and fulfill the need in several important ways:

(a)    by designing and implementing a high-fidelity snore signal acquisition system that reduces noise embedded in snore signals and adds standardization and consensus in the signal acquisition literature;

(b)    by devising an efficient preprocessing system that improves snore signal quality and detects snore activity;

(c)    by identifying effective OSA diagnostic markers that can reveal the UA acoustical features and nonlinear dynamical behavior, as well as correlate with AHI; and

(d)    by investigating physiological-anatomical-acoustical relationships of snores to validate the reliability of the diagnostic markers and to comprehend the mechanisms of snore production.

For any snore-based diagnostic methodologies, it is well-known that the OSA diagnostic accuracy depends on the acquired snore signal quality. Hence, it is essential to properly develop an acquisition system, which minimizes noise and ensures the integrity of data, even in a highly controlled setting of a sleep laboratory. The next chapter of this thesis discusses a robust system design and implementation for snore signal acquisition.

Chapter 3

# Snore Signal Acquisition System

The quality of snore signals plays a vital role in achieving accurate diagnostic results. Despite of a well-controlled sleep laboratory setting, the recorded snore signals are usually contaminated by background acoustical noise and electromagnetic interference (EMI) [102], lowering the diagnostic accuracy. Therefore, it is crucial to appropriately set up a snore signal acquisition system that reduces extraneous noise and interference.

This chapter describes the development of a robust system for acquiring snore signals in a clinical environment; the system minimizes noise pickup and interference, while providing user-friendly graphical interface with analytical tools for real-time signal evaluation. Details on the various functional elements in the acquisition system are systematically explained. Different aspects on the selection and installation of the system components (microphone, preamplifier, data acquisition card, grounding layout, and graphical user interface) are also discussed, with experiments to appraise the effectiveness of the proposed system in the reduction of background acoustical noise and EMI. Subsequently, we examined the effects of microphone position on snore signal quality and patient comfort, in terms of sleep laboratory impulse response analysis, signal quality evaluation, and patient feedback survey.

## 3.1    Instrumentation Setup

In this research work, we designed and implemented a high-fidelity and user-friendly acquisition system for snore signals in a sleep laboratory (Sleep Disorders Unit,

_____

The contents of this chapter are partly based on the author's publications [A6,A13].
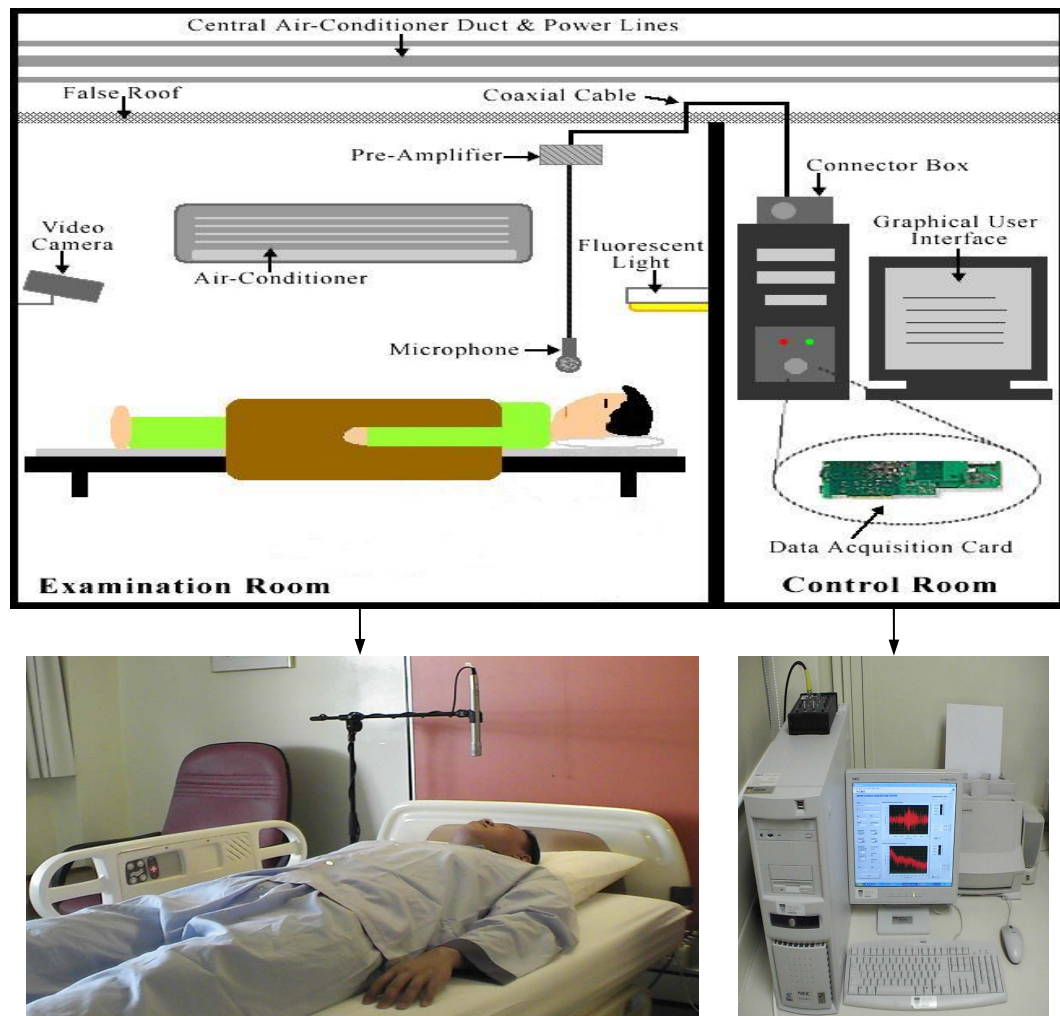
**Figure 3.1**    Snore signal acquisition system setup in a sleep laboratory.

Singapore General Hospital, Singapore), as pictured in Figure 3.1. The patient is bedded in the examination room, while the polysomnographic technicians are monitoring the acquisition process in the control room without interrupting the patient's sleep. The room setting gives the patient more comfort and privacy, which are essential for the reduction of first-night effect and to ensure an unbiased recording [103].

In the examination room, a non-invasive unidirectional condenser microphone (20–20000 Hz, model SM81, Shure Incorporated) is hung from a microphone boom stand about 0.3 m above the patient's mouth to capture snoring sounds. The

microphone is connected to a low-noise preamplifier (20–22000 Hz, model FP23, Shure Incorporated) with the output linked to a data acquisition card (model NI4552, National Instruments Corporation) in the control room through a double-shielded coaxial cable to minimize EMI during transmission. Snore signals are sampled at 44100 Hz with 16-bit resolution (compact disc quality) to record good quality snore signals and other sleep sounds.

## 3.2    Design Considerations

In the following, an overview of the various functional elements in the acquisition system is provided, before detailing the features of each system component. The measurand (sound) is picked up by a sensor (microphone) to produce an output, which is then converted from one energy form (pressure) to another (voltage) by a transducer. The transducer normally generates small electrical signals that require an additional signal conditioning unit (preamplifier) to boost its strength before transmission, so as to reduce any extraneous noise pickup and improve signal-to-noise ratio (SNR) at the output. Subsequently, the pre-amplified signals are transmitted via a double-shielded coaxial cable and converted to its equivalent digital form by an analog-to-digital converter (data acquisition card) for data storing and processing. Eventually, these data are analyzed and displayed in a form perceivable by users on a graphical user interface.

### 3.2.1    Microphone

The microphone is the first sensing element for snoring sounds. It transduces sound pressure variations into relative variations of electrical signal with a correct polar pattern, dynamic range, frequency response, and installation.

The microphone implemented in the acquisition system has a unidirectional

polar pattern rather than an omni- or a bi-directional pattern. A unidirectional microphone (pressure-gradient operated) has the highest response to sounds arriving from the front, with smoothly declining response to zero at the rear and sides of the microphone. This polar pattern enables most snoring sounds to be acquired while reducing background acoustical noise (e.g., sound from air conditioner and/or patient's movements) embedded in the snore signals. In contrast, an omni-directional microphone (pressure operated) has a uniform response to sounds from all directions, whereas a bi-directional microphone (pressure-gradient operated) mainly responses to sounds from the front and rear of the microphone [104].

Within a microphone, there is a transducer (dynamic or condenser type) that converts acoustic energy into electrical energy [104]. A dynamic microphone relies on the principle of electromagnetic induction, which the output is proportional to the velocity of a moving element within a magnetic field.  Conversely, a condenser microphone detects sounds using an electrically charged diaphragm, which forms a sound-sensitive capacitor. A condenser microphone is typically more expensive and complex than a dynamic microphone because it consists of an additional active circuitry that requires batteries or phantom power to operate. Nevertheless, a condenser microphone is widely used for vocal recording due to its high sensitivity and ability to deliver a broad range of tonal variations with a dynamic range more than 120 dB. Moreover, a condenser microphone has an extended frequency range with flat frequency response and an excellent transient response.

Snoring sounds, created by soft tissue vibrations and/or turbulent airflow at the narrowed UA, has a wide dynamic range of more than 90 dBA and a frequency range of 0–5000 Hz that fully characterizes the sound spectrum of snoring [99,105]. In order to acquire snore signals of high integrity, we chose a unidirectional condenser

microphone.

The unidirectional condenser microphone (20–20000 Hz, model SM81, Shure Incorporated) is positioned roughly 0.3 m away from the patient's mouth, where snoring sounds of low intensity can be captured, and patient can toss and turn comfortably without hitting the microphone (see Section 3.4 regarding the determination of mouth-to-microphone distance). Unlike a trachea microphone commonly used in PSG, the placement of a hanging microphone allows non-contact measurement, hence avoiding the use of lead wires around the patient and lessening the possibility of electrical shock hazards from contact. Thus, the proposed acquisition approach can increase patient safety and comfort during signal acquisition, reducing biased readings and achieving more reliable diagnostic results. In addition, an air-coupled microphone can better detect snore signals of frequency up to 4000 Hz, whereas a body-contact microphone is insensitivity to frequencies above 1000 Hz [106].

## 3.2.2    Preamplifier

The basic function of a preamplifier is to intensify the input differential signals while rejecting any existing common-mode signals. This is important especially during shallow breathing, where snoring sounds are usually soft and are easily distorted by stray noise during transmission. The suggested preamplifier (20–22000 Hz, model FP23, Shure Incorporated) for the acquisition system has low noise and a large common-mode rejection ratio of approximately 100 dB to diminish the effect of common-mode voltages. Furthermore, it has a selectable gain setting, which facilitates optimum amplification gain without signal clipping. Similar to the microphone, the preamplifier has a constant frequency response within the audio range of 20–20000 Hz.

Besides snore signals, somniloquous speech can also be clearly collected to optimize the classification of snore signals.

The preamplifier is installed in the examination room before the long coaxial cable, as rendered in Figure 3.1. This arrangement is supported and justified by the Friis' formula [107], popularly adopted in telecommunications engineering to compute noise factor (i.e., a measure of SNR degradation) consisting of several cascaded stages. Consider the preamplifier $p$ and the cable $c$ as two cascaded stages, having noise factor $F$ and gain $G$ denoted as $(F_p, G_p)$ and $(F_c, G_c)$, respectively. Since the low-noise preamplifier has a higher gain than the cable, it is reasonable to assume that $F_p < F_c$, and $G_p > G_c$. Therefore, based on the Friis' formula

$$F = F_x + \frac{F_y - 1}{G_x} \tag{3.1}$$

where $F$ is the overall noise factor of two cascaded stages, $x$ placed before $y$, the overall noise factor is expected to be smaller when the preamplifier is sited in the examination room before the cable, as compared to positioning the amplifier in the control room after the cable, i.e.,

$$\left( F_p + \frac{F_c - 1}{G_p} \right) \quad < \quad \left( F_c + \frac{F_p - 1}{G_c} \right). \tag{3.2}$$

### 3.2.3    Data Acquisition Card

The data acquisition card (model NI4552, National Instruments Corporation) implemented in this acquisition system has the ability to register signals with a bandwidth up to 95000 Hz, which entirely covers the bandwidth of snore signals. Moreover, the data acquisition card is configured to sample the input signals at 44100 Hz with 16-bit resolution. This configuration provides an analog-to-digital conversion
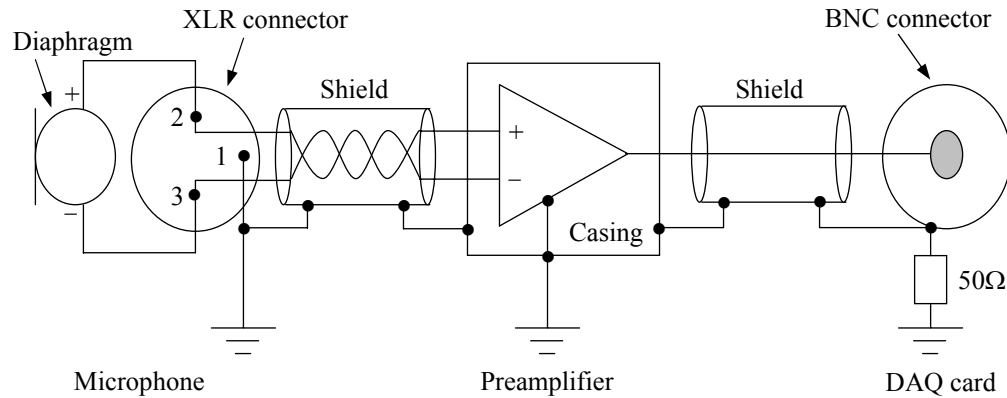
**Figure 3.2**   Circuit diagram of snore signal acquisition system.

with a good precision of 305.18 µV for the following optimal selections: amplification

gain = 66 dB and data acquisition card input range = ±10 V. Under these settings, most

acquired signals are found to utilize the input range without any clipping.

### 3.2.4    Grounding Layout

A circuit diagram of the snore signal acquisition system is illustrated in Figure 3.2. It

is recognized that one of the major undesirable disturbances when recording snore

signals is the EMI, originated from the proximity of the power lines and the electrical

or electronic devices via electromagnetic radiation [108-110]. Thus, it is advised to

reduce EMI and harmful signal pickup to an insignificant level through the use of

proper grounding and shielding techniques. Grounding offers reference for signals and

safety for patients and clinicians, whereas shielding minimizes the effects of disruptive

charge dissipation.

As can be seen in Figure 3.2, signals from the microphone are tapped to form a

balanced output, which is applied to the preamplifier operating in a differential mode

that responds to the voltage difference between the signals but ignores the common

noise embedded in both the signals. Connections between the microphone and the

preamplifier are made with a double-shielded coaxial cable to lower signal interference from external sources, yielding greater capacity to carry larger amount of undistorted signals. Furthermore, both signal and return wires are twisted and kept away from any potential magnetic-field regions to minimize electromagnetic coupling and extraneous cross-talk between neighboring wire pairs.

At the preamplifier, the output signals are fed into the data acquisition card, through a shielded coaxial cable, in an unbalanced form. These signals are not easily influenced by the EMI since it is largely amplified by a gain of 66 dB. Nonetheless, they are subjected to various fluctuations of the ground signals attributed to its single-ended configuration. To alleviate this problem, we integrated a pseudo-differential input circuit [111], comprising of a $50\Omega$ resistor between the negative input and the ground, into the circuit. The pseudo-differential input is analogous to a differential input that minimizes ground noise; however, unlike a differential input, the negative input does not change much with time but rather serves as a reference point for the signals at the positive terminal. With this, snore signal integrity is well-preserved even with an unbalanced output at the preamplifier.

### 3.2.5    Graphical User Interface

To allow easy communication between the polysomnographic technicians and the computers in the control room, as well as the instruments in the examination room, we developed a graphical user interface using LabVIEW$^{TM}$ (version 7.1, National Instruments Corporation). The user-friendly graphical user interface enables the technicians to effortlessly control the parameters of the instruments by means of interesting graphics models in the computer, with minimum typing. The graphical user interface design is based on the clinical documentation of patient biodata for PSG,
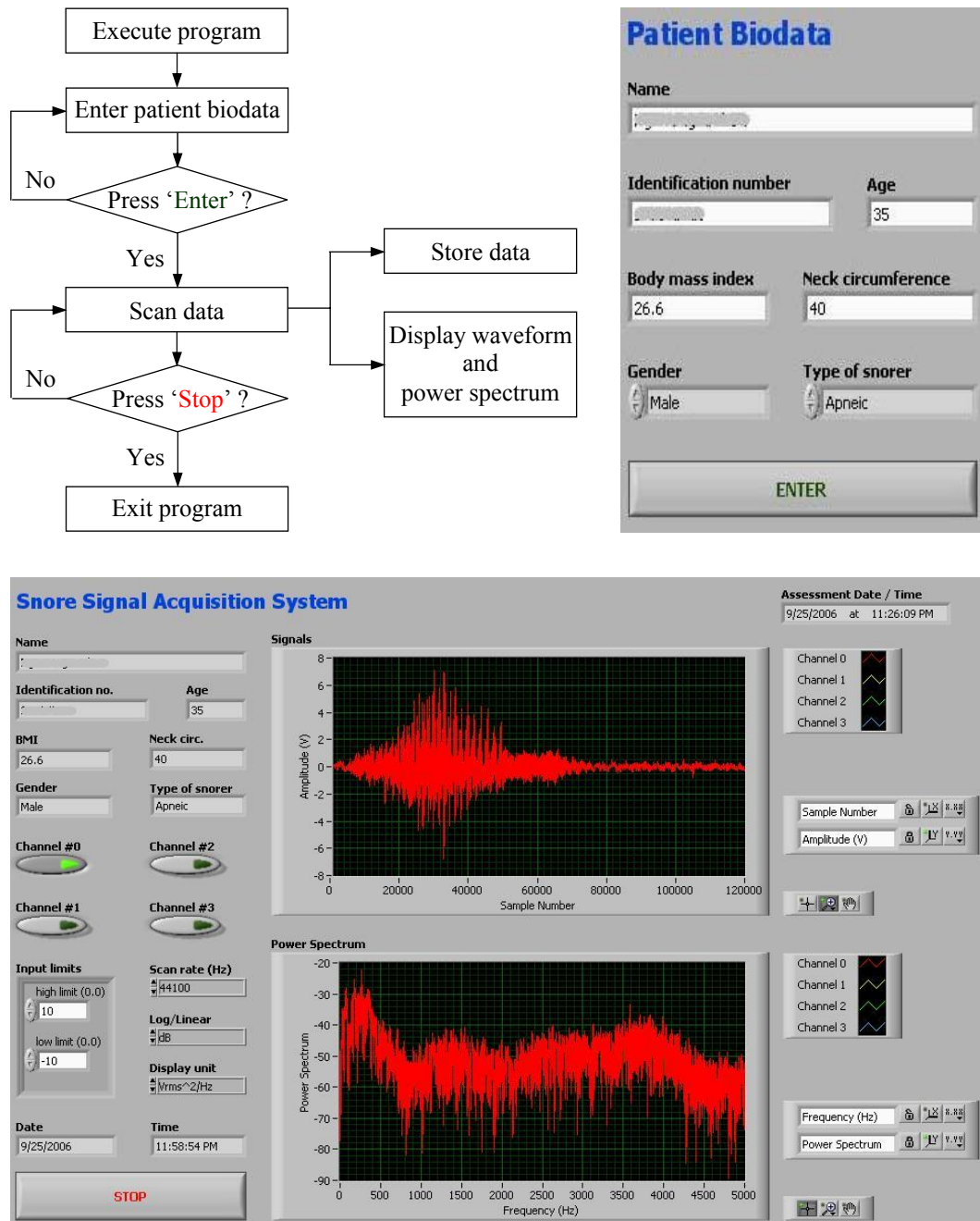
**Figure 3.3**     Graphical user interface of snore signal acquisition system including a program flowchart (top left), a pop-up window (top right), and a main window (bottom).

together with additional analytical tools for evaluating snore signals, as presented in Figure 3.3.

The pop-up window allows the technicians to enter the patient biodata, such as name, identification number, gender, age, body mass index, neck circumference, as
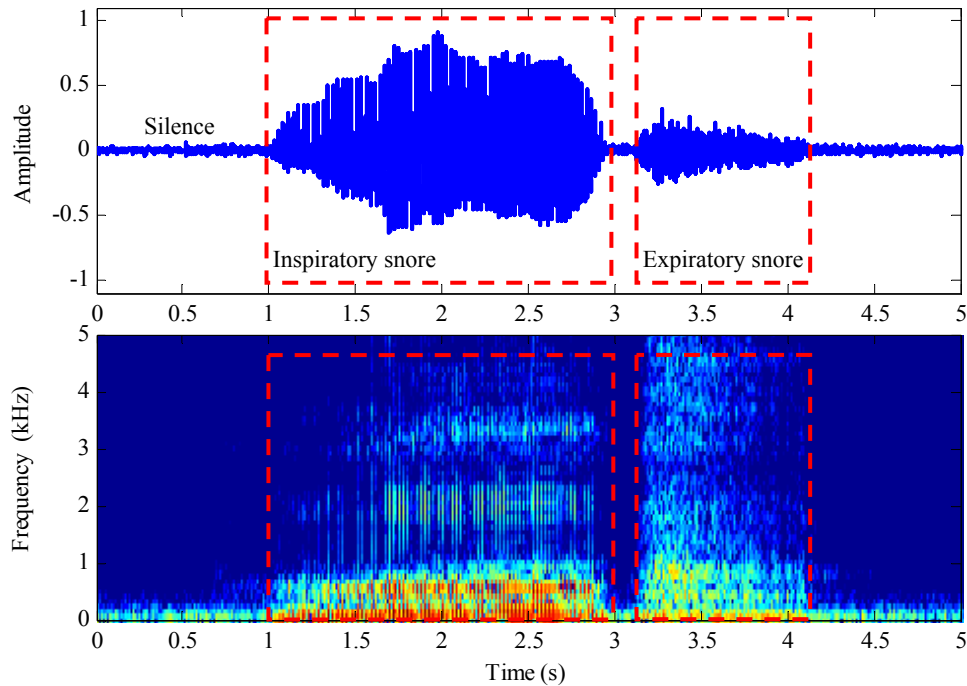
**Figure 3.4**    Temporal and spectral features of a snoring episode comprising of an inspiratory snore, an expiratory snore, and a silence segment.

well as type of snorer (apneic or benign) based on the patient's clinical history. In contrast, the front panel displays the patient biodata, real-time display of acquired signals and its corresponding power spectrum, along with selectable buttons and controllers for signal acquisition and analysis.

## 3.3    Acquisition System Appraisal

To appraise the robustness of the proposed acquisition system and assure that high quality snore signals are collected, two experiments were carried out to comparatively study the outcome of background acoustical noise and EMI before and after the implementation of the system. In order to perform a fair experiment, a segment of pre-recorded snoring sounds, which comprises of 10 similar snoring episodes with clear audible vibratory modes at the inspiratory phases, lasting about 60 s, was played back

via a speaker that simulated the patient's mouth, in the setup depicted in Figure 3.1. In this research work, we defined a snoring episode as an inspiratory and an expiratory snore segment with a short silence between them. Figure 3.4 shows both temporal and spectral features of a snoring episode. To add more realism, the experiments were conducted in the night when PSG usually commences.

### 3.3.1    Background Acoustical Noise

The effect of background acoustical noise was investigated with the preamplifier at different amplification gains, namely 50 dB, 54 dB, 58 dB, 62 dB, and 66 dB. For each gain setting, power spectral density (PSD) of the detected inspiratory snore signals was estimated using the Welch's averaged modified periodogram method [112,113], accompanied by a 1024-sample ($\approx$ 23 ms) Hanning window [113-115] with 75% overlap to avoid spectral leakage. As the name implies, the Welch's method computes PSD by averaging $M$ periodogram estimates, i.e.,

$$\hat{S}(f) = \frac{1}{M} \sum_{m=1}^{M} \hat{P}_m(f) \tag{3.3}$$

where

$$\hat{P}_m(f) = \frac{\Delta t}{N w_{\mathrm{P}}} \left| \sum_{k=1}^{N} x_m(k) w(k) \exp(-j 2\pi f k \Delta t) \right|^2 \tag{3.4}$$

is the periodogram of $m$th data segment with signal $x_m$ containing $N$ samples. $\Delta t$ represents the sampling period, $j = \sqrt{-1}$, and

$$w_{\mathrm{P}} = \frac{1}{N} \sum_{k=1}^{N} |w(k)|^2 \tag{3.5}$$

is the power of the temporal window $w(k)$, which removes the windowing effect from the total signal power in order to lessen the variance of the estimated PSD, providing
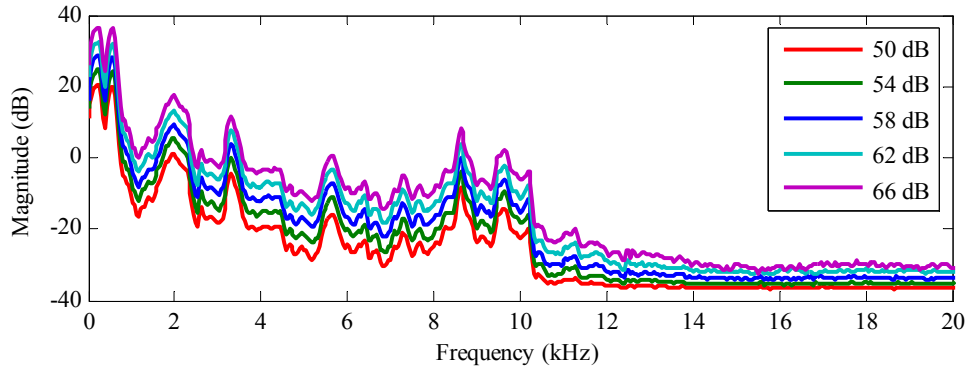
**Figure 3.5**   Power spectral densities of snore signals under different amplification gains: 50 dB, 54 dB, 58 dB, 62 dB, and 66 dB.

more control over the PSD bias/resolution properties [112,113]. The Hanning window function, on the other hand, is equated as

$$w(k) = \begin{cases} \dfrac{1}{2}\left(1 - \cos\left(\dfrac{2\pi(k-1)}{N-1}\right)\right) & , \quad 1 \le k \le N \\ 0 & , \quad \text{otherwise} \end{cases} \tag{3.6}$$

and it is preferred over other window functions (e.g., rectangular, triangular, Hamming, Blackman, and Kaiser) because of its popularity, easy implementation, and relatively good sidelobe behavior (-31 dB sidelobe level) [113-115].

In addition, SNR of the snore was also calculated. Due to the difficulties in obtaining the power of snore signals and the power of additive noise separately in the recording, the SNR was approximated by measuring the snore signal power $P_{signal+noise}$ during inspiratory snoring period and the noise power $P_{noise}$ during noise-only period. Since the snore signal and noise processes are independent,

$$\text{SNR} = \frac{P_{signal+noise} - P_{noise}}{P_{noise}}. \tag{3.7}$$

Figure 3.5 illustrates the PSD of the detected snore signals under various amplifications: 50 dB, 54 dB, 58 dB, 62 dB, and 66 dB, with their respective mean and standard deviation SNR = 8.12 ± 0.27 dB, 8.81 ± 0.48 dB, 9.31 ± 0.34 dB, 9.72 ± 0.25

dB, and 10.46 ± 0.44 dB. These results demonstrate that the preamplifier plays a key part in the acquisition system. SNR increases with increasing amplification gains, implying that the amplified signals exhibit clearer snoring sounds as they are less likely to be distorted by background acoustical noise. Moreover, frequency contents are stronger for signals of higher amplification, favoring the masking of ambient noise. Correspondingly, snoring sounds with high intensity level are not easily affected by noise pollution relative to snoring sounds of low intensity.

### 3.3.2     Electromagnetic Interference

The influence of the suggested grounding and shielding measures taken for EMI was examined in the similar setup highlighted in Figure 3.1, with the same segment of pre-recorded snoring sounds under the amplification gain of 66 dB for three different conditions: (a) before grounding and shielding; (b) after grounding and shielding with fluorescent lights on; and (c) after grounding and shielding with fluorescent lights off. PSD of the captured signals under these three conditions were computed and compared.

PSD of the captured snore signals before and after implementation of the suggested grounding and shielding measures, are rendered in Figure 3.6. Magnitudes of the PSD are very similar in the lower frequency range for all three conditions. However, at about 12.5 kHz, 15.5 kHz, and 18.8 kHz, unwanted spikes are spotted under the ungrounded and unshielded condition. Signals acquired under this condition are directly exposed to the electromagnetic radiation emitted from the central air conditioner duct and power lines above the false roof, leading to signal corruption and errors. Undesired harmonics and intermodulation products of the power lines may also add to the existence of the spike. On the contrary, under good grounding and harness
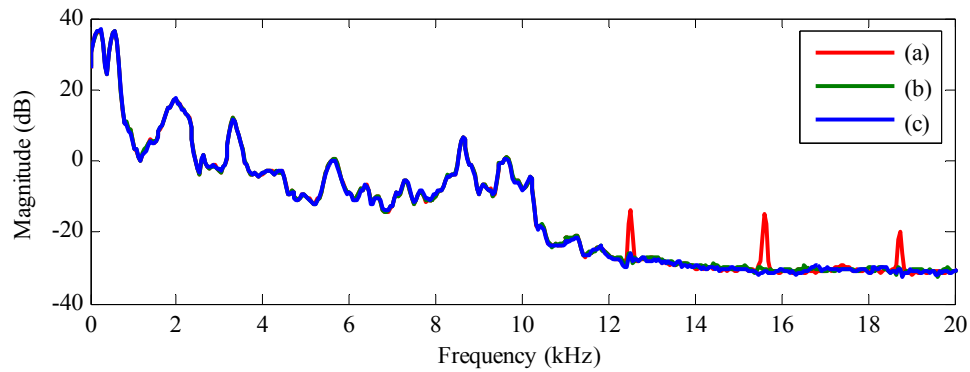
**Figure 3.6**     Power spectral densities of snore signals under different grounding and shielding conditions: (a) before grounding and shielding, (b) after grounding and shielding with fluorescent lights on, and (c) after grounding and shielding with fluorescent lights off.

shielding techniques in instrumentation, the high-frequency spikes are suppressed even when the fluorescent lights are on.


## 3.4     Mouth-to-Microphone Distance Determination

Apart from background acoustical noise and EMI, mouth-to-microphone distance (MMD, i.e., distance between the patient's mouth and the microphone) plays an important role in the quality of snore signals, together with the level of human comfort and safety. When MMD is too short, snore signals of good quality can be simply collected, but the patient may feel uneasy and find it hard to fall asleep. Conversely, when MMD is too long, it undoubtedly boosts the human's level of comfort and safety while at the expense of degrading signal intelligibility owing to increased pickup of surrounding noise and interference. We attempted to examine the impact of microphone positioning on snore signal quality and patient comfort by means of sleep laboratory impulse response analysis, signal quality evaluation, and patient feedback survey. In doing so, we hope to determine the appropriate MMD that can establish a good tradeoff between signal quality and patient comfort, refining the standards and

consensus for the measurement of snore signals.

### 3.4.1     Sleep Laboratory Impulse Response Analysis

Impulse response modeling is widely studied for room acoustical design [90]; it provides a reference calibration for an enclosed environment, allowing one to recreate or even optimize the auralization of the original impulse. The room impulse response was modeled using the image-source method attributable to its remarkable accuracy in identifying reflection paths [116,117]. The impulse response can be a function of sound source location $\mathbf{r}_s = [x_s\ y_s\ z_s]$, receiver or microphone location $\mathbf{r} = [x\ y\ z]$, room dimensions (length $L_x$, width $L_y$, and height $L_z$), and reflection coefficients of the six walls ($\beta_{x1}, \beta_{x2}, \beta_{y1}, \beta_{y2}, \beta_{z1},$ and $\beta_{z2}$). Explicitly, it can be expressed as

$$h\left(\mathbf{r},\mathbf{r}_s,t\right)=\sum_p\sum_u \beta_{x_1}^{|n-q|}\beta_{x_2}^{|n|}\beta_{y_1}^{|l-j|}\beta_{y_2}^{|l|}\beta_{z_1}^{|m-k|}\beta_{z_2}^{|m|}\frac{\delta(t-\tau)}{4\pi d} \tag{3.8}$$

where
$$\tau = \frac{\left\|\mathbf{R}_u+\mathbf{R}_p\right\|}{c} \tag{3.9}$$

denotes the arrival time of the reflected sound ray relative to any image source traveling at a velocity of $c \approx 345$ m/s in a room with temperature at 23ºC, and

$$d = \left\|\mathbf{R}_u+\mathbf{R}_p\right\| \tag{3.10}$$

defines the distance between any image source and the microphone. The vector

$$\mathbf{R}_p = \left[x_s - x + 2qx \quad y_s - y + 2jy \quad z_s - z + 2kz\right] \tag{3.11}$$

links the microphone location to the positions of the images at the first level. There are eight images at the first level, as each element of $p = (q, j, k)$ can hold the value 0 or 1. For higher image levels, the vector $\mathbf{R}_u = 2[nL_x\ lL_y\ mL_z]$ is added to $\mathbf{R}_p$, where each element of $u = (n, l, m)$ takes values between $-N$ and $+N$, depending on the number of image levels one prefers. Consequently, $\mathbf{R}_u$ delivers $(2N + 1)^3$ possible

combinations, giving a total of $8(2N+1)^3$ different paths.

To accurately simulate the received echo arrival time even at low sampling frequency, we replaced each impulse in Equation (3.8) by a low-pass-filtered impulse generated by the Hanning-windowed ideal low-pass filter [118]

$$w(t) = \begin{cases} \dfrac{1}{2}\left(1 + \cos\left(\dfrac{2\pi t}{t_{\mathrm{W}}}\right)\right)\mathrm{sinc}(2\pi f_{\mathrm{C}} t) \quad , & -\dfrac{t_{\mathrm{W}}}{2} < t < \dfrac{t_{\mathrm{W}}}{2} \\ 0 \quad , & \text{otherwise} \end{cases} \qquad (3.12)$$

where $t_{\mathrm{W}}$ is the window width, and $f_{\mathrm{C}}$ is the filter cutoff frequency. As advised in [116], we set $t_{\mathrm{W}} = 8$ ms, and $f_{\mathrm{C}} =$ Nyquist frequency of snore signals sampled at 44100 Hz.

Besides persevering the true echo arrival time, reverberation time $\mathrm{RT}_{60}$, which defines the time required for sound to decay by 60 dB, is another main factor to model a room impulse response. The $\mathrm{RT}_{60}$ relies on the volume of the room $V$, surface area $S_i$, and reflection coefficient $\beta_i$ of the $i$th wall, given by [90]

$$\mathrm{RT}_{60} = \frac{24\log_{\mathrm{e}}(10)V}{c\displaystyle\sum_{i=1}^{6} S_i\left(1 - \beta_i^2\right)} . \qquad (3.13)$$

Putting the sleep laboratory dimensions ($L_x = 4.74$ m, $L_y = 3.44$ m, and $L_z = 2.67$ m) and walls composition, for instance, smooth concrete, gypsum board, and light drapery, along with their respective absorption coefficients [90] $= 0.08$–$0.10$, $0.09$–$0.29$, and $0.03$–$0.35$, into Equation (3.13), the average $\mathrm{RT}_{60} \approx 1.0838$ s at 125–4000 Hz.

### 3.4.2     Signal Quality Evaluation

To evaluate the quality of snore signals in relation to the MMD, we collected snoring sounds in the similar setup depicted in Figure 3.1, with the same segment of pre-recorded snoring sounds under various MMDs beginning from 0.1 to 0.5 m at 0.1 m intervals. A digital sound level meter (30–130 dBA, Mastech MS6701, Precision

Mastech Enterprises Company) was mounted beside the microphone to quantify mean sound pressure level of each inspiratory snore.

The collected signals were processed within MATLAB$^{TM}$ and a digital audio software package (Cool Edit Pro$^{TM}$, version 1.2, Adobe Systems Incorporated). Similar to the earlier experiments for background acoustical noise and EMI, we estimated the PSD, SNR, and average root-mean-square power of the recorded snores. To facilitate easy comparison and discussion, we averaged each of the three evaluation metrics (sound pressure level, SNR, and root-mean-square power) for every microphone position.

### 3.4.3    Patient Feedback Survey

Apart from the objective measures, subjective analysis is equally essential, because if the patient does not feel comfortable throughout the night of signal acquisition, the recording can be severely biased, and eventually it may lead to inaccurate diagnostic results.

Feedbacks were obtained from the 40 patients listed in Table 1.1. In the sleep laboratory, as shown in Figure 3.1, we requested the patient to lie in the bed, raised the microphone distance at 0.1 m intervals starting from 0.1 m, and asked the patient to determine the minimum acceptable distance of the microphone from his or her mouth, with the consideration of possible tossing and turning during sleep.

### 3.4.4    Results and Discussion

Figure 3.7 illustrates the finite impulse responses of the sleep laboratory for different microphone positions. A close inspection of the first impulse of every position agrees with its corresponding distance. For example, the first impulse of 0.1 m is at a delay of
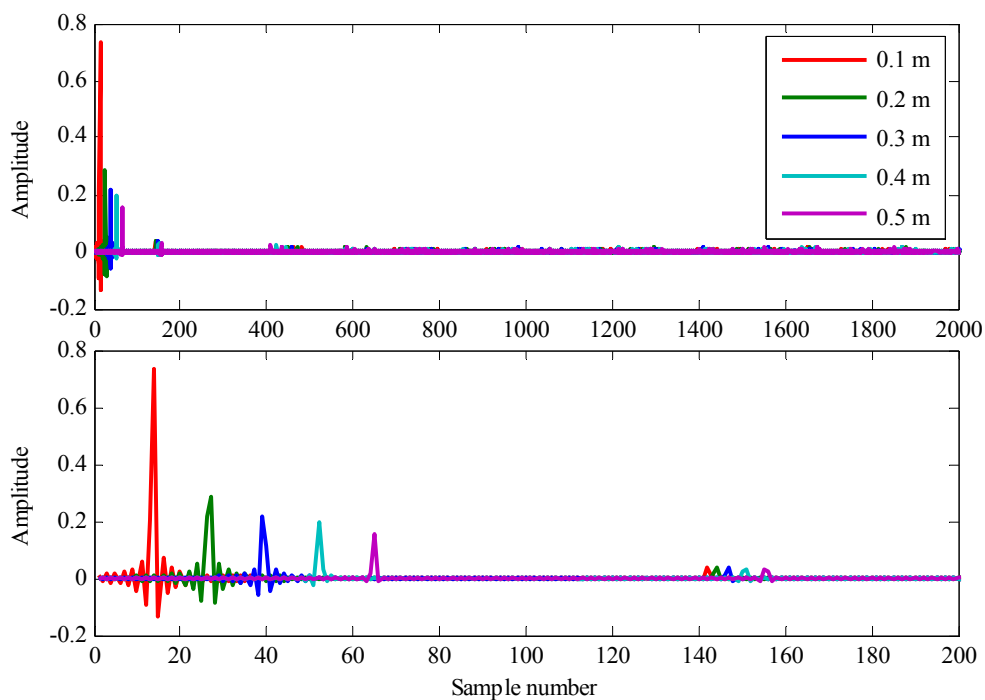
**Figure 3.7**   Impulse responses of a sleep laboratory for various mouth-to-microphone distances: 0.1 m, 0.2 m, 0.3 m, 0.4 m, and 0.5 m.
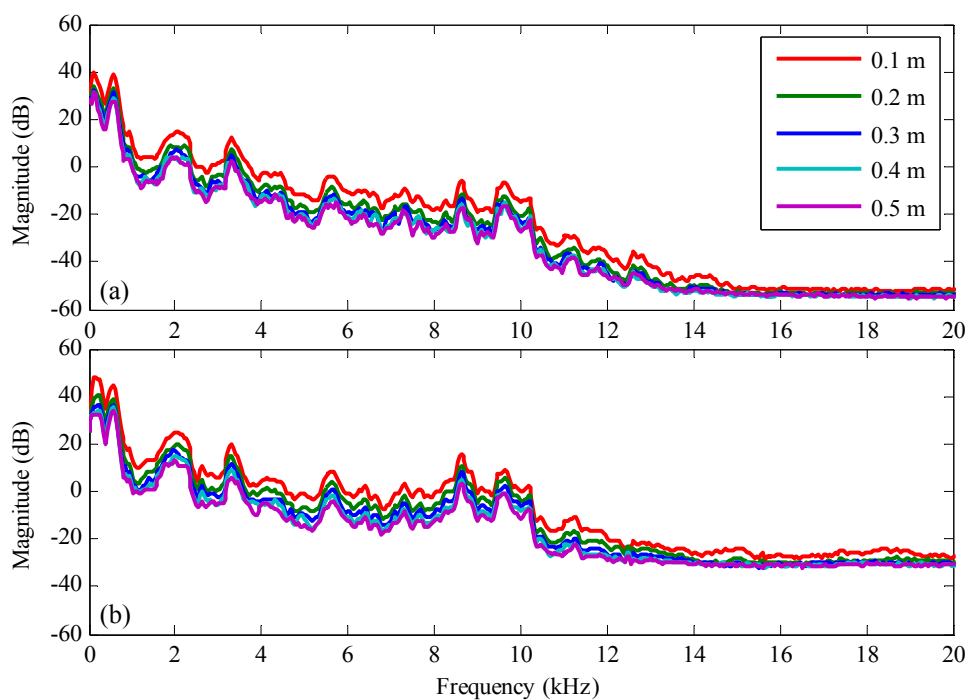


**Figure 3.8**   Power spectral densities of (a) simulated reverberant snore signals, and (b) on-site acquired snore signals for various mouth-to-microphone distances: 0.1 m, 0.2 m, 0.3 m, 0.4 m, and 0.5 m.

**Table 3.1**   Quality of snore signals for different mouth-to-microphone distances.

| MMD (m) | SPL (dBA) | | SNR (dB) | | RMS power (dBFS) | |
|---------|-----------|-----------|----------|----------|------------------|-----------|
| 0.1 | 75.09 | (75.09) | 20.27 | (21.91) | -13.06 | (-20.86) |
| 0.2 | 68.80 | (69.07) | 13.54 | (17.22) | -19.93 | (-25.73) |
| 0.3 | 66.66 | (65.55) | 10.46 | (15.67) | -23.07 | (-27.46) |
| 0.4 | 65.05 | (63.05) | 7.79 | (14.99) | -24.92 | (-28.23) |
| 0.5 | 64.36 | (61.11) | 7.14 | (14.71) | -25.92 | (-28.56) |

Values are presented as mean. Values within brackets were computed from the inverse distance law for SPL, and from reverberant snores for SNR and RMS power. MMD refers to mouth-to-microphone distance in metres; SPL, sound pressure level in A-weighted decibels; SNR, signal-to-noise ratio in decibels; RMS power, root-mean-square power in decibels relative to full-scale sine wave.

14 samples ($\approx$ 0.3175 ms), which is the traveling time taken by sound in a direct path. Furthermore, the strength of the impulses gradually declines with increasing distance and time, yielding a decreased power magnitude of the reverberant snoring sounds produced by convolving the segment of snoring sounds with the sleep laboratory impulse response, as clearly displayed in Figure 3.8a.

The attenuation of signal power is also noted in Figure 3.8b, where PSD of the on-site acquired snore signals are plotted. Inasmuch as the simulated reverberant snores and the on-site acquired snores exhibit the same spectral dynamics below 5000 Hz (snore maximum frequency-of-interest [99,105]), it is obvious that the latter ones contain undesired noise and artifacts due to the nature of non-invasive sound measurement, being more severe for MMD = 0.5 m.

Interestingly, we observed that short MMD does not always amount to good signal quality; instead, it enhances the likelihood of triggering a proximity effect and capturing explosive breath sounds, especially when the patient recovers from an OSA attack. The blast of explosive air slams into the microphone diaphragm, causing overload and impairing the sound intelligibility, as evidenced by the presence of strong
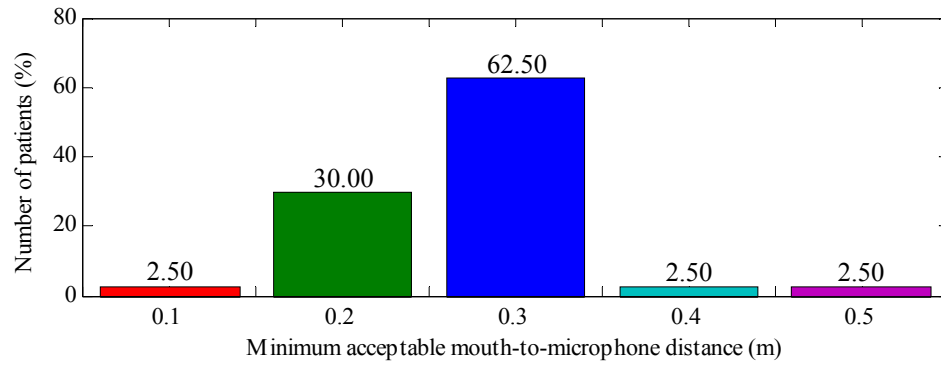
**Figure 3.9**   Patient feedback score for different mouth-to-microphone distances: 0.1 m, 0.2 m, 0.3 m, 0.4 m, and 0.5 m.

fluctuating high-frequency modes for MMD = 0.1 m in Figure 3.8b.

Table 3.1 summarizes the quality of the captured snores by means of sound pressure level, SNR, and root-mean-square power. We comparatively evaluated the quantitative results with those computed from the inverse distance law [90] for sound pressure level, as well as from reverberant snores for SNR and root-mean-square power, marked within brackets. Results consistently indicate that the intelligibility of snore signals degrades with increasing MMD. When MMD rises from 0.1 to 0.4 m, the three evaluation metrics drop more than 10 dB (sound pressure level from 75.09 to 65.05 dBA; SNR from 20.27 to 7.79 dB; and root-mean-square power from -13.06 to -24.92 dBFS).

On the contrary, the patient feedback score, plotted in Figure 3.9, apparently highlights that majority (62.5%) of the patients feel comfortable and secure when MMD = 0.3 m, followed by 30.0% when MMD = 0.2 m. As compared to 0.2 m, the MMD = 0.3 m enables the patients to toss and turn without any worries of hitting the microphone during sleep. Such worries are certainly understandable, especially for patients with OSA as they frequently experience repetitive transient arousals from sleep [8-10].

Taking into account the signal quality and human comfort and safety, we recommended a MMD of 0.3 m. This value of MMD is also put forward by the Union of European Phoniatricians for measuring the human voice (e.g., speaking and singing) [119].

## 3.5    Summary

This chapter discusses the design and implementation of a high-fidelity and user-friendly system for acquisition of snore signals in a sleep laboratory, highlighting various considerations for the selection and installation of system components (microphone, preamplifier, data acquisition card, grounding layout, and graphical user interface). Experiments were performed to warrant the robustness of the proposed acquisition system by investigating the effect of background acoustical noise, EMI, and MMD. Results reveal (a) a reduction in background acoustical noise after appropriate selection of the amplification gain; (b) the removal of high-frequency spikes associated with EMI after proper grounding and shielding; (c) a decrement in signal intelligibility and an increment in human comfort with increasing MMD; and (d) a microphone setting of 0.3 m from the patient's mouth can achieve a good tradeoff between signal quality and patient comfort.

Besides developing a robust snore signal acquisition system to minimize unwanted noise and EMI, the non-contact nature of data measurement is often susceptible to external noise distortion, which must be eliminated prior to signal analysis and interpretation. The next chapter of this thesis devises a novel wavelet-driven preprocessing system that further improves snore signal quality and detects snore activity.

Chapter 4

# Snore Signal Preprocessing System

Appropriate instrumentation and measurement techniques are essential for achieving snore signals of high integrity while lessening patient discomfort. Nonetheless, owing to the fact that snore signals are acquired via a non-invasive microphone, additive background acoustical noise can inevitably degrade signal fidelity, reducing SNR and, most importantly, yielding inaccurate diagnostic results. Thus, it is always vital to eliminate and discern snore signals from any undesired noise before subsequent analysis and classification. Unfortunately, such issues on snore signal enhancement and snore activity (SA) detection in a clinical environment have not been specifically addressed in the literature (see Chapter 2).

This chapter proposes an efficient preprocessing system for snore signals, which comprises of a level-correlation-dependent (LCD) threshold to suppress background acoustical noise, and a SA detector to simultaneously identify snore segment boundaries in a translation-invariant discrete wavelet transform (TIDWT) domain. Environmental condition of a sleep laboratory is highlighted, focusing on the characteristics of environmental noise (i.e., background acoustical noise embedded in snore signals). Subsequently, we provided a detailed overview of the LCD threshold and SA detector in the proposed preprocessing system, as well as a brief discussion on the comparison methods and evaluation metrics. Experimental conditions, including the dataset for analysis and the selection procedures for optimal parameters, are also outlined. Finally, experimental results, in terms of theoretical statistics quality, SNR,

---

The contents of this chapter are partly based on the author's publications [A4,A10,A12].

mean opinion score (MOS), and clinical usefulness of the system, are presented and discussed.

## 4.1    Snore Signal Enhancement

### 4.1.1    Environmental Noise Properties

In order to understand the inherent noise in snore signals and develop a robust noise suppression algorithm, we studied the environmental noise in the sleep laboratory where the acquisition system for snore signals is implemented.

Based on the setup rendered in Figure 3.1, we comparatively evaluated the spectral behavior of the noise captured with that of fractal noises (e.g., white, pink, and brown). Through observations and consensus among polysomnographic technicians in the Sleep Disorders Unit, the air conditioner in the laboratory is the most possible source of noise. Hence, we recorded the environmental noise via the unidirectional microphone sited above the patient bed, without the patient's presence, for the three available air conditioner settings (low, medium, and high). We further modeled the power spectral density $S(f)$ of the noise as a function, which is inversely proportional to the signal frequency with an exponent $\beta$,

$$S(f) \propto \frac{1}{f^{\beta}} \tag{4.1}$$

where $\beta = 0$, 1, and 2, for white, pink, and brown noise, respectively [120].

In the acquired environmental noise, we noted that $\beta \approx 1.9801$, 2.0520, and 2.1069 for the low, medium, and high air conditioner setting, correspondingly, as illustrated in Figure 4.1. Since $\beta$ is constantly close to 2, one can infer that the background acoustical noise embedded in the captured snore signals is colored, and wavelet-like bases are the best functions to represent these signals as the role of
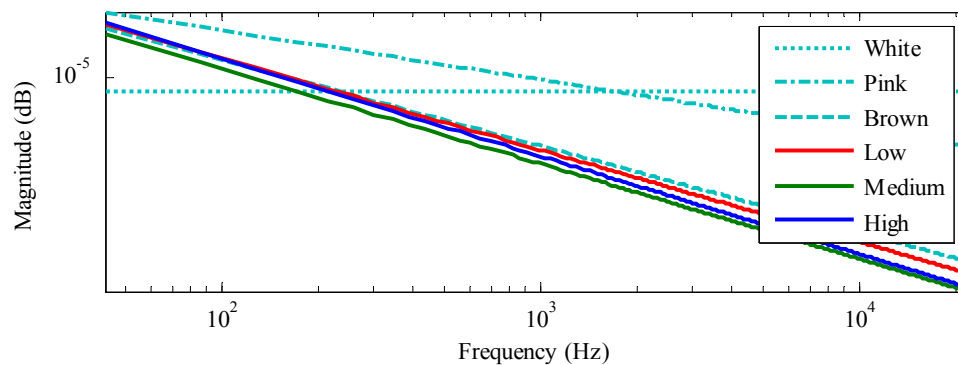
**Figure 4.1** Power spectral density log-log slopes for various noise types: white, pink, brown, and environmental noise for low, medium, and high air conditioner settings.

wavelet basis is a Karhunen-Loeve-type expansion for $1/f$ processes [121,122].

### 4.1.2    Translation-Invariant Discrete Wavelet Transform

During the last decade, wavelet thresholding for denoising signals has evolved extensively because of its capability to optimize both temporal and spectral information, thereby better preserving the desired signals while removing unwanted noise. Two initial considerations for a wavelet thresholding scheme include the type of wavelet transform and the choice of denoising threshold.

A TIDWT is a non-orthogonal, undecimated variant of the classical (orthogonal, maximally-decimated) discrete wavelet transform (DWT). The key motivation behind developing TIDWT is to overcome the visual artifacts attributable to translation variance of DWT, for instance, pseudo-Gibbs phenomena in the neighborhood of discontinuities [123,124]. Two typical schemes for TIDWT implementation are the à-trous and the cycle-spinning [125,126]. The à-trous scheme, in essence, relinquishes downsampling operator from DWT but upsamples filter responses. Conversely, the cycle-spinning scheme applies DWT to all circular shifts of an $N$-element input vector and averages over all these shifts. Due to periodic time-varying property of the rate-

change operators in the decomposition filter banks, the number of operations for cycle-spinning scheme reduces from $O(N^2)$ to $O(N \log_2 N)$ [124,127]. In other words, instead of performing all $N$ circular shifts, the output is now computed for the original input and its shift-by-one. Nevertheless, for both schemes, TIDWT yields redundant coefficients that offer momentous advantages in noise reduction, for example, better root mean squared error properties (10–20%) than classical DWT denoising [123,124]. In this work, we considered the cycle-spinning scheme owing to its popularity and flexibility [125].

Figure 4.2 depicts a block diagram of the proposed snore signal preprocessing system in TIDWT analysis/synthesis filterbank for two scale levels. Let $\mathbf{x} = \{x_k\}_{k=1}^{N}$ be a finite-length desired snore signal that is contaminated by an additive colored Gaussian noise $\mathbf{z} = \{z_k\}_{k=1}^{N}$ with a standard deviation $\sigma$. The resulting noisy snore signal at the receiving end is $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z} = \{y_k\}_{k=1}^{N}$. Under TIDWT, the transformed noisy signal becomes $\mathbf{Wy} = \mathbf{Wx} + \sigma\mathbf{Wz}$, where

$$\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_j, \ldots, \mathbf{W}_J, \mathbf{W}_{J+1}]^T \tag{4.2}$$

is an $(J+1)N \times N$ matrix, and $J$ is the maximal level of decomposition. Each $\mathbf{W}_j$ is an $N \times N$ sub-matrix whose columns are circularly shifted versions of the impulse response of wavelet filters $\mathbf{g} = \{g_k : k \in \mathbf{Z}\}$ that produces wavelet coefficients at scale level $j$, or likewise, scaling filters $\mathbf{h} = \{h_k : k \in \mathbf{Z}\}$ that produces scaling coefficients for $\mathbf{W}_{J+1}$. Since the scaling filters and wavelet filters are normalized,

$$\mathrm{diag}\left(\mathbf{W}_j^T \mathbf{W}_j\right) = 1 \in \mathfrak{R}^N, \qquad \forall j \in [1 \, ; J+1] \tag{4.3}$$

perfect reconstruction $\mathbf{MW} = \mathbf{I}$ can be attained, where

$$\mathbf{M} = [\frac{1}{2}\mathbf{W}_1, \ldots, \frac{1}{2^j}\mathbf{W}_j, \ldots, \frac{1}{2^J}\mathbf{W}_J, \frac{1}{2^J}\mathbf{W}_{J+1}] \tag{4.4}$$
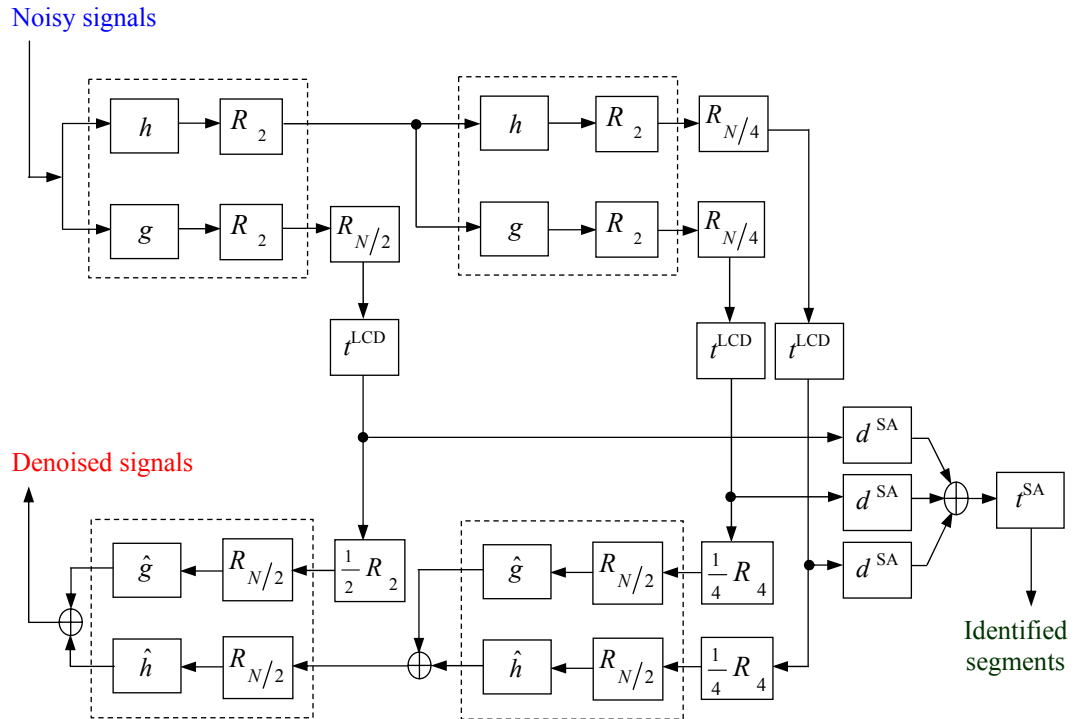
**Figure 4.2**   Snore signal preprocessing system in translation-invariant discrete wavelet transform analysis/synthesis filterbank for two scale levels. $\hat{h}$ and $\hat{g}$ denote the complex conjugate of the scaling filters $h$ and wavelet filters $g$, respectively. $t^{LCD}$ denotes level-correlation-dependent threshold, while $d^{SA}$ and $t^{SA}$ denote snore activity (SA) detector and SA threshold, respectively. $R_m$ denotes the sequence reordering operator with a list of elements of an $m \times N/m$ matrix by rows, where $N$ is the input vector length.

is the pseudoinverse matrix of the TIDWT, and the division by powers of two compensates for the increasing redundancy of the transformation [128]. Similar to TIDWT, the pseudoinverse TIDWT takes $O(N \log_2 N)$ operations [128,129].

As discussed earlier, the output for the even and the odd shifts in cycle-spinning scheme are obtained by computing DWT for the original input and its shift by one, respectively. For convenience, these recursive computational procedures can be simplified by introducing a sequence reordering operator $R_m$ [129], which competently arranges the coefficient sequences to substitute the downsampling operators without

affecting the redundancy, and therefore, the shift invariance. Consider an $N$-element input vector $\mathbf{y} = \{y_k\}_{k=1}^N$ and $m$ divides $N$, the sequence reordering operator

$$R_m(\mathbf{y}) = \left(y_1, y_{m+1}, \ldots, y_{N-(m-1)}, \ldots, y_m, y_{2m}, \ldots, y_N\right) \tag{4.5}$$

is a list of elements of an $m \times N/m$ matrix by rows. The inverse of $R_m$ is $R_{N/m}$,

$R_{N/m}(R_m(\mathbf{y})) = \mathbf{y}$.

### 4.1.3     Level-Correlation-Dependent Threshold

TIDWT denoising can be summarized in the following four steps: (a) performs TIDWT on the time-domain signal; (b) identifies an appropriate threshold; (c) implements a suitable thresholding rule to shrink or retain the effects of certain wavelet coefficients; and (d) inverts the resulting coefficients to reconstruct the denoised time-domain signal. Explicitly, the denoised signal can be given by

$$\hat{x}_k = \mathbf{M}\left\{\eta\left(w_{jk}, \sigma_j \lambda_j\right)\right\}, \qquad \forall j \in [1\,;J+1], \quad \forall k \in [1\,;N] \tag{4.6}$$

where $\eta$ can either denote a soft-thresholding rule [130]

$$\eta_S\left(w_{jk}, \sigma_j \lambda_j\right) = \mathrm{sgn}\left(w_{jk}\right)\left(\left|w_{jk}\right| - \sigma_j \lambda_j\right) \tag{4.7}$$

or a hard-thresholding rule [131]

$$\eta_H\left(w_{jk}, \sigma_j \lambda_j\right) = w_{jk} \cdot 1_{\left[\left|w_{jk}\right| \geq \sigma_j \lambda_j\right]} \tag{4.8}$$

or some compromise between the two rules. The terms $\sigma_j$ and $\lambda_j$ represent the noise variance estimate and the threshold at scale level $j$, correspondingly. To improve the finite sample properties of the threshold, we set $\lambda_{J+1} = 0$ [131,132].

Many thresholds have been proposed in the wavelet literature. Three popular ones are the VisuShrink universal threshold [131], the Stein's unbiased risk estimator [133], and the level-dependent (LD) threshold [134]. The former two thresholds are commonly implemented with the soft-thresholding rule for suppressing additive white

Gaussian noise, whereas the latter is for removing correlated noise. Although these thresholds achieve remarkable denoising performances (e.g., better convergence rate and optimal in the minimax sense [130]), there are two inherent shortcomings: (a) white noise is useful as a conceptual entity, but it seldom occurs in reality; and (b) these thresholds are tailored for DWT, which has lesser impact in denoising applications than TIDWT. Consequently, these thresholds are not suitable for this research work, where the acquired snore signals are corrupted by colored noise in the sleep laboratory environment, and the chosen wavelet transform is translation invariance. To address this concern, we proposed a LCD threshold that can optimally enhance snore signals in the presence of colored noise while providing translation invariance.

The LCD threshold

$$t_j^{\text{LCD}} = \sigma_j \lambda_j = \sigma_j \sqrt{2(1+\phi_j)\log_e((1+j)N)}, \qquad \forall j \in [1;J] \qquad (4.9)$$

consists of a correlation limit

$$\phi_j = \max_{k \neq l} |r_{kl}| < 1 \qquad (4.10)$$

where

$$r_{kl} = \text{Cov}(\xi_k, \xi_l) \qquad (4.11)$$

describes the correlation between random variables $\{\xi_k\}_{k=1}^{(1+j)N} = \mathbf{Wz}$ produced at scale level $j$, and a conditional factor $(1 + j)$ in the log-term that accounts for redundant information introduced at $j$ [135]. To further maximize noise suppression, the LCD threshold is accompanied with a noise variance estimator calculated from the median absolute deviation (MAD) of the wavelet coefficients at each scale level and a factor 0.6745 for calibration with the Gaussian distribution [134]

$$\sigma_j = \frac{\text{MAD}(w_{jk})}{0.6745}, \qquad \forall j \in [1;J], \quad \forall k \in [1;N]. \qquad (4.12)$$

The LCD threshold is formulated based on the properties of the TIDWT decomposition of the colored background acoustical noise. To demonstrate these properties, a length of $2^{12}$ points ($\approx$ 93 ms) was extracted from each of the three environmental signals mentioned in Section 4.1.1 and decomposed into 12 levels using TIDWT with Daubechies compactly supported filter of length 4 [136,137]. Results are tabulated in Table 4.1, with the following observations.

(a)     Coefficients in the same scale are mostly correlated, not independent. Within-scale (i.e., both are related with same scale level, $j$) correlation was analyzed via the normalized autocorrelation function [138]. To facilitate easy comparisons, we attempted to fit the normalized autocorrelation values verse the lag sample to a logarithmic function $c = a + b \log_{10}(s)$ that interprets as follows: $c$ is the squared of the auto-correlation value, $a$ is the correlation value at a unit lag sample $s = 1$, and $b$ is the declining rate in $s^{-1}$. The parameters $a$ and $b$ for each $j$ are listed in Table 4.1. For all the environmental noisy signals, results consistently reveal that the correlation values of coefficients within each scale not only increase but also fade away rapidly as we move towards the coarsest scale.

(b)     Coefficients between scales are generally correlated. The degree of correlation for between-scale (i.e., both are related with successive levels, $j$ and $j + 1$) can be expressed by the Spearman's rank correlation coefficient $r_s$ [139,140], a non-parametric measure of correlation. This statistical test is better than the standard correlation coefficient because the latter will only operate when variables are related in a linear relation. In real situations, assuming a linear relation is often unpractical. An alpha of 0.05 was used. For a significance level of p-value < 0.05, there is a

**Table 4.1**   Statistical properties of noise coefficients in detail scale levels of translation-invariant discrete wavelet transform for low, medium, and high air conditioner settings.

| Scale level | Low air conditioner setting | | | |
|:---:|:---:|:---:|:---:|:---:|
| $j$ | $a$ | $b$ | $r_s$ | p-value |
| 1 | 0.0181 | -0.0055 | 0.0280 | 0.0736 |
| 2 | 0.0185 | -0.0056 | -0.0105 | 0.5019 |
| 3 | 0.0415 | -0.0126 | -0.1228 | <0.0001 |
| 4 | 0.0685 | -0.0207 | -0.1352 | <0.0001 |
| 5 | 0.1246 | -0.0373 | -0.1996 | <0.0001 |
| 6 | 0.2093 | -0.0620 | -0.1338 | <0.0001 |
| 7 | 0.4200 | -0.1228 | -0.2778 | <0.0001 |
| 8 | 0.6119 | -0.1771 | -0.1467 | <0.0001 |
| 9 | 0.7546 | -0.2227 | 0.0031 | 0.8433 |
| 10 | 1.1340 | -0.3305 | -0.0061 | 0.6973 |
| 11 | 1.3887 | -0.3887 | 0.0090 | 0.5633 |
| 12 | 1.3891 | -0.3957 | - | - |

| Scale level | Medium air conditioner setting | | | | High air conditioner setting | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| $j$ | $a$ | $b$ | $r_s$ | p-value | $a$ | $b$ | $r_s$ | p-value |
| 1 | 0.0173 | -0.0053 | 0.0353 | 0.0237 | 0.0174 | -0.0053 | 0.0340 | 0.0296 |
| 2 | 0.0186 | -0.0057 | -0.0419 | 0.0073 | 0.0186 | -0.0057 | 0.0442 | 0.0047 |
| 3 | 0.0379 | -0.0115 | -0.1327 | <0.0001 | 0.0458 | -0.0139 | -0.1735 | <0.0001 |
| 4 | 0.0647 | -0.0196 | -0.1474 | <0.0001 | 0.0718 | -0.0216 | -0.0797 | <0.0001 |
| 5 | 0.1043 | -0.0314 | 0.0158 | 0.3132 | 0.1266 | -0.0380 | -0.1758 | <0.0001 |
| 6 | 0.2138 | -0.0636 | -0.2141 | <0.0001 | 0.1970 | -0.0591 | -0.0667 | <0.0001 |
| 7 | 0.3880 | -0.1149 | -0.2998 | <0.0001 | 0.3803 | -0.1131 | -0.3522 | <0.0001 |
| 8 | 0.5897 | -0.1736 | -0.0751 | <0.0001 | 0.7048 | -0.2066 | -0.0987 | <0.0001 |
| 9 | 0.7643 | -0.2260 | -0.0419 | 0.0073 | 0.8578 | -0.2484 | -0.0185 | 0.2375 |
| 10 | 1.1775 | -0.3383 | 0.0656 | <0.0001 | 1.3028 | -0.3674 | -0.0325 | 0.0377 |
| 11 | 1.3953 | -0.3923 | 0.0181 | 0.2473 | 1.2635 | -0.3579 | 0.0266 | 0.0884 |
| 12 | 1.3425 | -0.3835 | - | - | 1.5963 | -0.4524 | - | - |

$a$ refers to correlation value at a unit lag sample; $b$, declining rate in the logarithmic function $c = a + b \log_{10}(s)$ for within-scale correlation measure; $r_s$, Spearman's rank correlation coefficient; p-value, level of significance for between-scale ($j$ and $j+1$) correlation measure. A p-value < 0.05 was considered statistically significant.

95% chance that an association exists between the coefficients. Besides

that, the strength of an association can be signified by $|r_s|$. As $|r_s| \rightarrow 1$, the

strength of the monotonic relationship increases. However, when $r_s = 0$, no monotonic relationship is present. Both $r_s$ and p-value for each pair of scale levels ($j$ and $j+1$) are summarized in Table 4.1. These results advocate that the coefficients possess certain degree of correlation between scales, regardless of the air conditioner settings.

Based on the above two observations, one can confirm that correlation between noise coefficients exists in TIDWT, both within- and between-scale. The most likely explanation for the existence of correlation is the inherent spectral features of the noise and the non-orthonormality of TIDWT. This unique interaction property between coefficients and TIDWT is perhaps one of the contributions to the success of wavelet-domain motion estimation [141,142]. Here, for denoising applications, we incorporated a correlation limit in the LD threshold to handle this association characteristic between coefficients, with the aim to optimize noise cancellation in translation invariance domain.

To determine the correlation limit in the proposed LCD threshold, we examined all the possible correlations in the TIDWT domain. A number of formulas have been previously derived that highlighted the relationship between the covariance $r_{kl}$ in Equation (4.11) and the set of all translations and dilations in both the scaling function and wavelet function at dyadic rationals [135]. Among these formulas, the one that can best approximate the $r_{kl}$ is the correlation between scaling coefficients. To illustrate this, let $c_k^j$ be the scaling coefficients at scale $j$, and $\mathbf{z}^{\text{i.i.d.}} = \left\{ z_k^{\text{i.i.d.}} \right\}_{k=1}^{N}$ be the independent identically distributed (i.i.d.) random variables input. Then the covariances between $c_k^j$ satisfy the following formulas:

$$
\begin{aligned}
\mathrm{Cov}\left(c_k^j, c_{k+r}^j\right) &= \sum_{n,m} h_n h_m \mathrm{Cov}\left(c_{k+2^{j-1}n}^{j-1}, c_{k+r+2^{j-1}m}^{j-1}\right) \\
&= \sum_{n,m} h_n h_m \Phi\left(2^{-j+1}r + m - n\right) \\
&= \sum_t \Phi\left(2^{-j+1}r + t\right) \sum_m h_{m+t} h_m \\
&= \sum_t \Phi\left(2^{-j+1}r + t\right) a_t = \Phi\left(2^{-j}r\right), \qquad r \neq 0, \quad r \in Z.
\end{aligned}
$$

For $j = 1$,

$$
\begin{aligned}
\mathrm{Cov}\left(c_k^1, c_{k+r}^1\right) &= \sum_{n,m} h_n h_m \mathrm{Cov}\left(z_{k+n}^{\mathrm{i.i.d.}}, z_{k+r+m}^{\mathrm{i.i.d.}}\right) \\
&= \sum_m h_{m+r} h_m = a_r = \Phi\left(2^{-1}r\right)
\end{aligned}
$$

are the à-trous filter coefficients in the scaling equation [127] for the autocorrelation function $\Phi$ of the compactly supported scaling function.

The autocorrelation function $\Phi$ is the fundamental function of the symmetric iterative interpolation scheme referred in [143]. The outcome of repeating the Lagrange iterative interpolation process converges to the values of $\Phi$ at dyadic rationals as $j \to \infty$, and $\Phi(2^j) \to 1$ as $j \to \infty$. Analogously, when the same process is applied to the orthogonal Daubechies scaling function, an approximation of the scaling function values can be obtained [135]. Several researchers [127,144] have further explored the applications of $\Phi$ for compactly supported wavelets. They demonstrated that $\Phi(2^j)$ can possibly dominate the correlation terms of scaling and wavelet coefficients at scale level $j$, as well as the cross-correlation terms. For instance, the inequality in [144],

$$
\Psi(x) = 2\Phi(2x) - \Phi(x) \tag{4.13}
$$

suggests that $\Phi(2^j) > \Psi(2^j)$. Attributed to these properties of $\Phi$, one can presume that $\Phi(2^j)$ is the best estimate of the covariance $r_{kl}$ at $j$. Table 4.2 indicates the correlation limit in Equation (4.10) for different length of Daubechies filters at different $j$.

**Table 4.2**   Correlation limit for different length of Daubechies filters at different scale levels.

| | Scale level | | | | | |
|---|---|---|---|---|---|---|
| Daub | 1 | 2 | 3 | 4 | 5 | 6 |
| 2 | 0.5000 | 0.7500 | 0.8750 | 0.9375 | 0.9688 | 0.9844 |
| 4 | 0.5625 | 0.8438 | 0.9492 | 0.9844 | 0.9954 | 0.9987 |
| 6 | 0.5859 | 0.8709 | 0.9639 | 0.9904 | 0.9975 | 0.9993 |

| | Scale level | | | | | |
|---|---|---|---|---|---|---|
| Daub | 7 | 8 | 9 | 10 | 11 | 12 |
| 2 | 0.9922 | 0.9961 | 0.9980 | 0.9990 | 0.9995 | 0.9998 |
| 4 | 0.9996 | 0.9999 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 6 | 0.9998 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |

Daub refers to length of Daubechies filter.

As the noise variance for colored noise can change from one frequency band to another, the LCD threshold is not a constant but varies with scale levels owing to threshold dependency on the noise variance. The increasing LCD threshold is consistent with the phenomenon that denoising via TIDWT requires slightly larger threshold than denoising via DWT [135,145]. When the wavelet transform is orthogonal (i.e., DWT where correlation coefficients and redundant information vanish), the LCD threshold functions like the LD threshold [134]

$$t_j^{\text{LD}} = \sigma_j \sqrt{2(1+0)\log_e((1+0)N)} = \sigma_j \sqrt{2\log_e(N)}\,. \tag{4.14}$$

## 4.2    Snore Activity Detection

Apart from signal enhancement, SA detection (i.e., the process of discriminating snore presence and silence or background noise) is another essential preprocessing stage before analyzing the acoustical properties of snores to distinguish between patients with and without OSA. Several studies [98,146,147] on segmentation and classification of snores have been published recently. Conventional measures, such as

short-time energy and zero-crossing rate, were adopted in these studies to determine snore segment boundaries. While providing sufficient cues for boundary detection, these measures are based on averages over windows of fixed length and allow only a limited flexibility in the selection of the time-frequency resolution. Wavelet transform, on the other hand, are well-localized in both time and frequency. Coarse scale wavelets are localized in frequency, whereas detail scale wavelets are localized in time. As a consequence of this localization property of wavelet decomposition, the detection of signals through wavelet transform may be accomplished more accurately. Here, we proposed a SA detector via TIDWT to explore the feasibility of using TIDWT for SA detection.

### 4.2.1    Snore Activity Detector

As pointed out in Section 3.3, a snoring episode can be characterized as an inspiratory and an expiratory snore segment with a brief silence between them. Attributable to the background noise, the desired snore signals are contaminated and the noise spectrum profile can be estimated in the silence region. In the proposed preprocessing system, where snore signals are enhanced and SA are identified, the snoring episode is first decomposed into several scale levels using TIDWT.

As an illustrative example in Figure 4.2, the wavelet coefficients in each level are thresholded based on the LCD threshold, and subsequently manipulated within a SA detector to construct a SA envelope. Because of the multiresolution analysis property of wavelet transform, the envelope for each scale level provides useful cues for determining snore segments correctly. The envelopes for all levels, including the one formed by the scaling coefficients, are eventually summed up to obtain a dynamic SA feature for segment boundary detection. Figure 4.3 exemplifies SA detection for
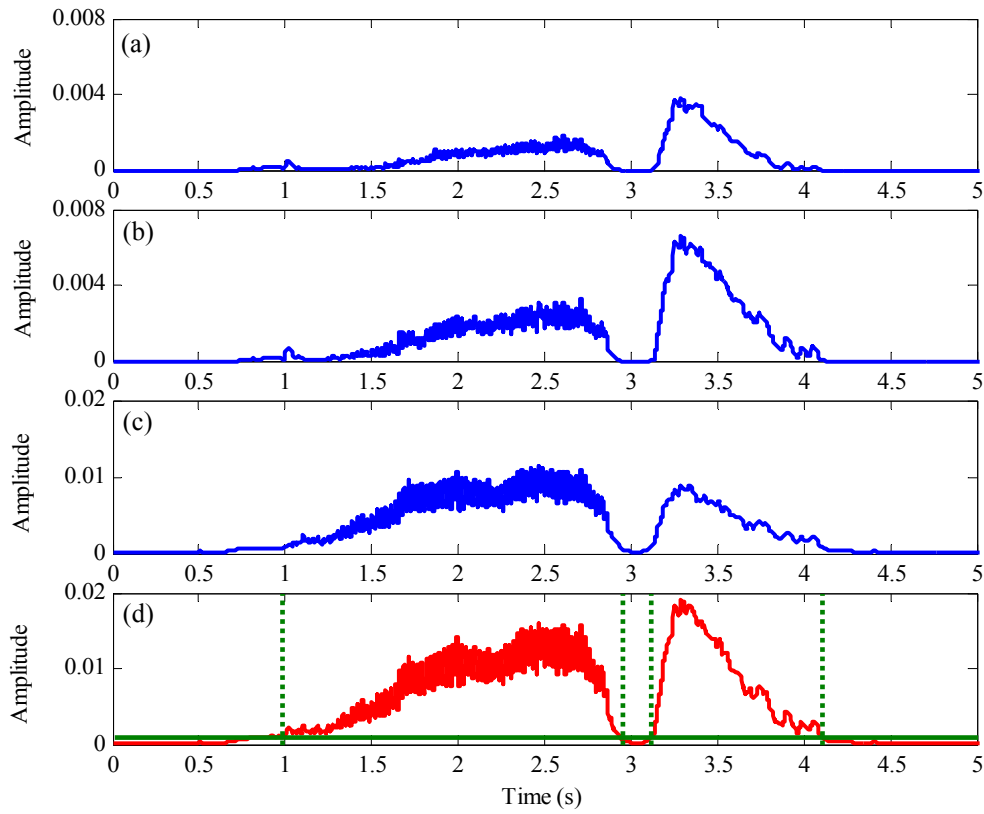
**Figure 4.3**   Envelopes for (a) detail scale level 1, (b) detail scale level 2, (c) coarse scale level 2, and (d) snore activity feature of a snoring episode undergoing snore activity detection, with snore activity threshold and segment boundaries indicated by green solid line and dotted lines, respectively.

two scale levels on the snoring episode shown in Figure 3.4. We noticed that the magnitudes of the SA feature in both the inspiratory and expiratory snore segments are always greater than those in the silence regions.

The SA detector has a working principle quite similar to the shifted delta cepstrum widely used in speech recognition [148]. It is the mean absolute value of the delta coefficients defined as

$$d_i^j = \frac{1}{K}\sum_{k=1}^{K}\left|\Delta c_{i,k}^j\right| \tag{4.15}$$

$$\text{where} \qquad \Delta c_{i,k}^{j} = \frac{\sum_{n=-N}^{N} nc_{i,k+n}^{j}}{\sum_{n=-N}^{N} n^2} \qquad\qquad (4.16)$$

denotes the $k$th delta coefficient in the $i$th frame at $j$th level. The delta coefficient is obtained by a first-order polynomial over a size of $(2N+1)$ coefficients, centered on the current coefficient. Unlike the delta cepstrum that uses cepstrum coefficients from neighboring frames (interframe processing), the SA detector uses coefficients within a particular frame (intraframe processing). In a nutshell, the SA detector measures the local variation of each coefficient and emphasizes the dynamic aspects of the snore signals spectrum.

Since snore signals are almost noise-free after wavelet thresholding, only a simple SA threshold is needed to establish the endpoints of snore segments. This threshold is derived from the statistics of the SA feature, i.e.,

$$t^{\text{SA}} = \mu_{\text{S}} + 3\sigma_{\text{S}} \qquad\qquad (4.17)$$

where $\mu_{\text{S}}$ and $\sigma_{\text{S}}$ are the mean and standard deviation within the silence regions, respectively. SA decision is made when the SA threshold is exceeded. Considering a Gaussian probability density function [139], the threshold can reject up to 99.7% of the silence samples, thus accepting mostly the snore signal samples.

## 4.3    Preprocessing System Appraisal

### 4.3.1    Experimental Conditions

#### 4.3.1.1    Patient Dataset

Snoring sounds of the 40 patients (30 apneic and 10 benign) tabulated in Table 1.1 were captured in parallel with PSG via the snore signal acquisition system (see Chapter 3) in the sleep laboratory. 10 snoring episodes of length $2^{18}$ points ($\approx 6$ s) for

each episode (equivalent to at least a complete breathing cycle) were randomly extracted from each patient with segment boundaries marked out by a group of 8 normal-hearing polysomnographic technicians from the Sleep Disorders Unit and 8 normal-hearing signal processing specialists from our institution using Cool Edit Pro$^{TM}$, giving a total of 400 snoring episodes for analysis.

### 4.3.1.2    Optimal Parameter Selection

At the core of any wavelet-based denoising method, apart from the type of wavelet transform and the thresholding methodology as discussed previously, there are other decisions to be made regarding the selection of the mother wavelet function and the optimal level of decomposition. The nature of the mother wavelet defines the functional basis which the signal of interest is decomposed. The more a wavelet resembles the signal, the better it denoises the signal. It is often desired that the wavelet has properties, such as regularity, vanishing moments, and symmetry. In this work, we chose the Daubechies family of wavelets due to their outstanding fundamental vanishing moment property, or equivalently, the capability of the scaling functions to reproduce maximal degree of polynomials with minimal support [136,137].

Within the Daubechies wavelet family, we adopted filter length 4 (Daub4) and its corresponding 17 decomposition levels for snore signal enhancement via the LCD threshold. Figure 4.4 renders the average SNR across the existing dataset for each possible length of Daubechies filter and level of decomposition.  It can be clearly seen that Daub4 with 17 decomposition levels delivers higher SNR improvement than the other possibilities. A probable reason is that as a lower order wavelet, it can better replicate the fast variations in snore signals. Furthermore, the hard-thresholding rule
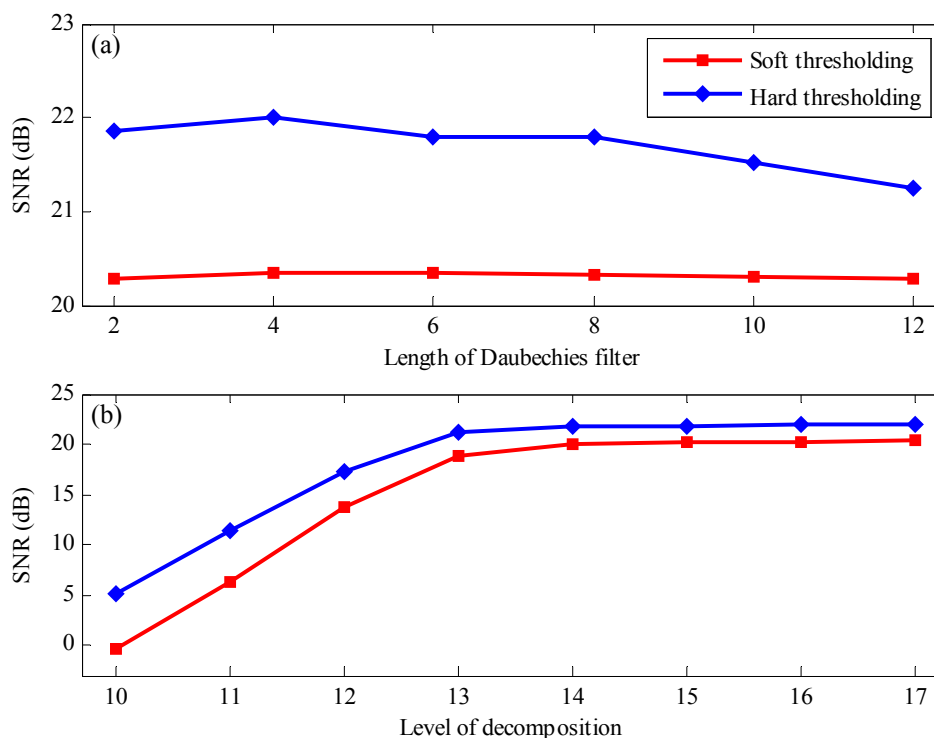
**Figure 4.4** Signal-to-noise ratio (SNR) for (a) different length of Daubechies filters at decomposition level = 17 and (b) different level of decomposition at filter length = 4, under soft- and hard-thresholding rules.
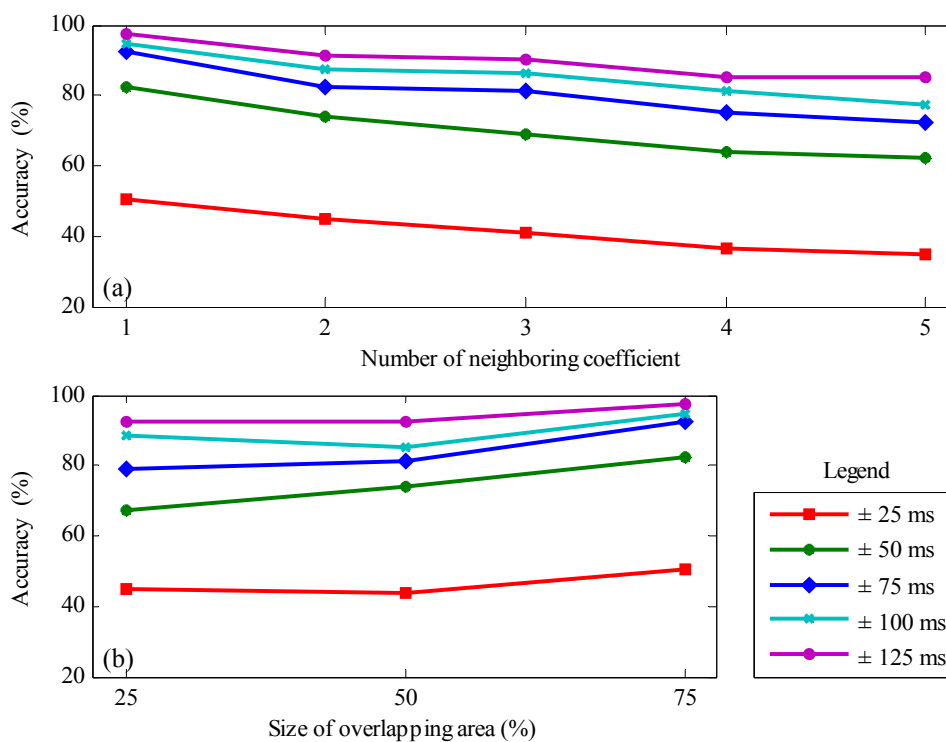


**Figure 4.5** Snore segment boundary detection accuracy for (a) different neighboring coefficients at overlapping size = 75% and (b) different overlapping sizes at neighboring coefficient = 1, given by several tolerance degrees: ±25 ms, ±50 ms, ±75 ms, ±100 ms, and ±125 ms.

outperforms the soft-thresholding rule.

Immediately after thresholding, the wavelet-based SA detector operates on a frame-by-frame basis at each scale level $j$. The preferred frame size is 1024 sample ($\approx$ 23 ms) because the usual fundamental frequency of snore signals can be as low as 50 Hz ($\approx$ 20 ms) [25,149,150]; hence, snore signals are likely to be quasi-stationary in such a time period, as in speech signal processing [151,152]. To prevent smearing, the successive frames were 75% overlapped using a Hanning window [113-115], equated in Equation (3.6). Moreover, the intraframe neighboring coefficient $N = 1$ is recommended as it can yield a higher accuracy for boundary detection relative to the other possibilities. The justifications for selecting these parameters are revealed in Figure 4.5, which highlights the results for detecting snore segment boundaries given by the various tolerance degrees ($\pm25$ ms, $\pm50$ ms, $\pm75$ ms, $\pm100$ ms, and $\pm125$ ms) through the SA detector and for different overlapping sizes and neighboring coefficients.

### 4.3.2     Snore Signal Enhancement

#### 4.3.2.1     Theoretical Statistics

Theoretical investigations were carried out to assess the effectiveness of the proposed LCD threshold based on statistical theorems mentioned in [135,153]. Assuming that the noise coefficients in TIDWT domain follow a Gaussian distribution (a property that is preserved during wavelet transformation [154]), then the difference between the probability of correlated random variables below a threshold and the probability of independent identically distributed random variables below the same threshold can be expressed as

$$\left| P\left\{ \max_{k=1,\dots N} \left| \xi_k \right| \leq \lambda \right\} - P\left\{ \max_{k=1,\dots N} \left| z_k^{\text{i.i.d.}} \right| \leq \lambda \right\} \right| \leq K \sum_{1 \leq k < l \leq N} \left| r_{kl} \right| \exp\left( -\frac{\lambda^2}{1 + \left| r_{kl} \right|} \right) \qquad (4.18)$$

with an error term and a constant $K$. The objective is to determine a threshold

$$\lambda \geq \sqrt{2 \log_e ((1+j)N)}, \qquad \forall j \in [1;J] \qquad (4.19)$$

to account for the maximal size of random variables at level $j$, and that the error term

$$\sum_{1 \leq k < l \leq (1+j)N} \left| r_{kl} \right| \exp\left( -\frac{\lambda^2}{1 + \left| r_{kl} \right|} \right) \to 0 \qquad \text{as} \qquad N \to \infty. \qquad (4.20)$$

By fulfilling the previous criteria, it also implies that $P\left\{ \max \left| z_k^{\text{i.i.d.}} \right| \leq \lambda \right\}_{k=1}^{(1+j)N} \to 1$ and

$P\left\{ \max \left| \xi_k \right| \leq \lambda \right\}_{k=1}^{(1+j)N} \to 1$ as $N \to \infty$, a similar analytical approach used for deriving the

universal threshold with soft-thresholding rule [130,131]. A possible threshold that

fulfills the previous criteria is the proposed LCD threshold in Equation (4.9). A proof

is presented in the following lemma.

Consider the maximal level of decomposition $j = J$, $r_{kl} = \text{Cov}(\xi_k, \xi_l)$, where

$\left\{ \xi_k \right\}_{k=1}^{(1+J)N} = \mathbf{Wz}$ is the correlated random variables with zero mean and a unit standard

deviation, as well as $\phi = \max_{k \neq l} \left| r_{kl} \right| < 1$. The proposed LCD threshold

$$t^{\text{LCD}} = \sqrt{2(1+\phi) \log_e ((1+J)N)}$$

can ideally achieve the error term

$$\sum_{1 \leq k < l \leq (1+J)N} \left| r_{kl} \right| \exp\left( -\frac{\lambda^2}{1 + \left| r_{kl} \right|} \right) \to 0 \qquad \text{as} \qquad N \to \infty.$$

*Proof*: Owing to the orthogonal properties of scaling and wavelet filters, several

number of covariance $r_{kl}$ are zero [135]. The total number of nonzero terms for $1 \leq k <$

$l \leq (1+J)N$ can be approximated as follows:

$$
\begin{aligned}
\#\{r_{kl} \mid r_{kl} \neq 0\} \;=\; & \#\{\mathrm{Cov}(c_m^J, c_n^J) \mid \mathrm{Cov}(c_m^J, c_n^J) \neq 0\} \\
& + \#\{\mathrm{Cov}(c_m^J, d_n^j) \mid \mathrm{Cov}(c_m^J, d_n^j) \neq 0, \qquad \forall j \in [1\,;J]\} \\
& + \#\{\mathrm{Cov}(d_m^j, d_n^i) \mid \mathrm{Cov}(d_m^j, d_n^i) \neq 0, \qquad \forall j \in [1\,;J], \quad \forall i \in [1\,;j]\} \\
\leq \;& DN(2^J - 1) + DN2^J \sum_{n=1}^{J} \frac{2^n - 1}{2^n} + DN \sum_{j=1}^{J} 2^j \sum_{n=1}^{j} \frac{2^n - 1}{2^n} \\
\leq \;& DN(2^J - 1 + J2^J + J2^{1+J} - 2^{1+J} + 2) \\[6pt]
\leq \;& DN(2^J(3J - 1) + 1)
\end{aligned}
$$

where $D$ = (length of Daubechies compactly supported filter)/2. Accordingly, the error term gives

$$
\begin{aligned}
\sum_{1 \leq k < l \leq (1+J)N} |r_{kl}| \exp\!\left(-\frac{(t^{\mathrm{LCD}})^2}{1 + |r_{kl}|}\right) \;=\; & \sum_{1 \leq k < l \leq (1+J)N} |r_{kl}| \exp\!\left(-2\log_e\big((1+J)N\big)\frac{1+\phi}{1+|r_{kl}|}\right) \\
\leq \;& \sum_{1 \leq k < l \leq (1+J)N} |r_{kl}| \exp\!\big(-2\log_e\big((1+J)N\big)\big) \\
\leq \;& \frac{DN(2^J(3J-1)+1)}{\big((1+J)N\big)^2}.
\end{aligned}
$$

For $J < \log_2 N$,

$$
\sum_{1 \leq k < l \leq (1+J)N} |r_{kl}| \exp\!\left(-\frac{(t^{\mathrm{LCD}})^2}{1 + |r_{kl}|}\right) \;\leq\; \frac{D(2^J(3J-1)+1)}{N(1+J)^2} \to 0 \qquad \text{as} \qquad N \to \infty.
$$

For $J = \log_2 N$,

$$
\sum_{1 \leq k < l \leq (1+J)N} |r_{kl}| \exp\!\left(-\frac{(t^{\mathrm{LCD}})^2}{1 + |r_{kl}|}\right) \;\leq\; \frac{D(N(3\log_2 N - 1)+1)}{(1+\log_2 N)^2 \, N} \to 0 \qquad \text{as} \qquad N \to \infty.
$$

The correlation limit plays a significant role in the LCD threshold, especially when $J = \log_2 N$ as shown next. If $\phi = 0$, then

$$
t^{\mathrm{LCD}} = \sqrt{2\log_e\big((1+J)N\big)}
$$

and

$$\sum_{1 \le k < l \le (1+J)N} |r_{kl}| \exp\left(-\frac{\left(t^{\text{LCD}}\right)^2}{1+|r_{kl}|}\right) = \sum_{1 \le k < l \le (1+J)N} |r_{kl}| \exp\left(-\frac{2\log_e\left((1+J)N\right)}{1+|r_{kl}|}\right)$$

$$\le \frac{DN\left(2^J(3J-1)+1\right)}{\left((1+J)N\right)^{\frac{2}{1+|r_{kl}|}}}.$$

For $J < \log_2 N$,

$$\sum_{1 \le k < l \le (1+J)N} |r_{kl}| \exp\left(-\frac{\left(t^{\text{LCD}}\right)^2}{1+|r_{kl}|}\right) \le \frac{D\left(2^J(3J-1)+1\right)}{\left(1+J\right)^{\frac{2}{1+|r_{kl}|}} N^{\frac{1-|r_{kl}|}{1+|r_{kl}|}}} \to 0 \qquad \text{as} \qquad N \to \infty.$$

The LCD threshold without the correlation limit still works well for $J < \log_2 N$. On the contrary, it may not function properly for $J = \log_2 N$ as the error term does not approach zero as $N \to \infty$, i.e.,

$$\sum_{1 \le k < l \le (1+J)N} |r_{kl}| \exp\left(-\frac{\left(t^{\text{LCD}}\right)^2}{1+|r_{kl}|}\right) \le \frac{DN\left(N(3\log_2 N - 1)+1\right)}{\left(1+\log_2 N\right)^{\frac{2}{1+|r_{kl}|}} N^{\frac{2}{1+|r_{kl}|}}}$$

$$\le \frac{DN^{\frac{2|r_{kl}|}{1+|r_{kl}|}}(3\log_2 N - 1)}{\left(1+\log_2 N\right)^{\frac{2}{1+|r_{kl}|}}}$$

$$+ \frac{D}{\left(1+\log_2 N\right)^{\frac{2}{1+|r_{kl}|}} N^{\frac{1-|r_{kl}|}{1+|r_{kl}|}}} \to \infty \qquad \text{as} \qquad N \to \infty.$$

From the previous derivation, one may reasonably infer that the proposed LCD threshold is suitable for denoising application in TIDWT domain for all cases. However, to avoid overthresholding, we utilized the LCD threshold at two critical scale levels: (a) the finest level of decomposition ($j = 1$), where most noise coefficients are present with minimum useful signal coefficients; and (b) the optimal level of decomposition ($j = J$) to guarantee that error term approaches zero. At the remaining scale levels ($1 < j < J$), a smaller LCD threshold can be employed by setting $\phi = 0$.
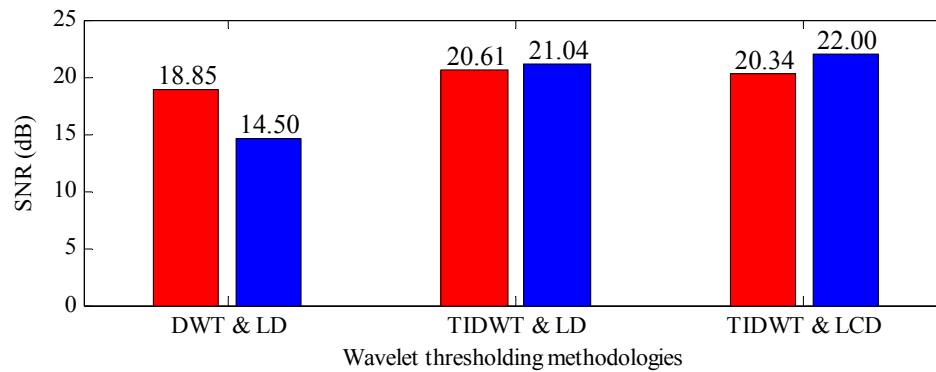
**Figure 4.6** Signal-to-noise ratio (SNR) comparison for various wavelet thresholding methodologies with soft thresholding (red) and hard thresholding (blue): discrete wavelet transform (DWT) and level-dependent (LD) threshold, translation-invariant discrete wavelet transform (TIDWT) and LD threshold, and TIDWT and level-correlation-dependent (LCD) threshold.

### 4.3.2.2   Objective Measures

Besides the theoretical assessment, the LCD threshold was comparatively evaluated with the LD threshold under the soft- and the hard-thresholding rules for DWT and TIDWT by means of SNR equated in Equation (3.7).

The improvement in average SNR of the denoised snoring episodes is illustrated in Figure 4.6. Results are provided as net improvement so that relative effectiveness can be observed from the six comparison methodologies, namely LD threshold with DWT, LD threshold with TIDWT, and LCD threshold with TIDWT, for both soft- and hard-thresholding rules. The proposed LCD threshold under the hard-thresholding rule performs much better (0.96–7.50 dB) than the other methodologies, with an SNR improvement of 3.02–38.22 dB over noisy snore signals. Conversely, when the LCD threshold is applied with the soft-thresholding rule, the SNR improvement is slightly smaller (0.27–0.70 dB) than that of the LD threshold for both thresholding rules, which may be attributed to oversmoothing of the denoised signals, causing unnecessary reduction in signal amplitude. These results are in agreement with the

earlier discussions. Firstly, translation-invariant denoising is found to be superior to classical DWT in denoising. Secondly, because of redundancy and correlation between coefficients, translation-invariant denoising requires slightly higher thresholds than orthogonal DWT denoising. Nonetheless, careful implementation of the threshold for TIDWT denoising is crucial to alleviate the effects of overthresholding. Finally, hard thresholding may be more appropriate for translation-invariant denoising than soft thresholding.

### 4.3.2.3    Subjective Measures

A subjective perceptual MOS measure [151,152] was utilized to augment the objective SNR measure for accessing the naturalness and intelligibility of the enhanced snore signals. MOS was yielded by having the group of polysomnographic technicians and signal processing specialists rate the quality of the enhanced snore signals on a 5-point scale where 5 indicates excellent quality and 1 indicates bad quality with annoying and objectionable distortion. The scores were then averaged, giving a single evaluation metric for each comparison methodology.

Subjective outcomes using MOS for the same comparison methodologies are depicted in Figure 4.7. The quality rating ranges from 1 (bad) to 5 (excellent). There is an interesting difference between the MOS and SNR results. While the relative MOS for hard thresholding are in line with the corresponding SNR value, soft thresholding for LCD threshold receives higher rating than LD threshold in TIDWT. The net result is that the LD threshold with soft thresholding did not perform as well as the LCD threshold with soft thresholding for TIDWT. This difference between the MOS and SNR results may be a consequence of the presence or absence of short-term spurious oscillations in the denoised signals that are known to have stronger influences in one's
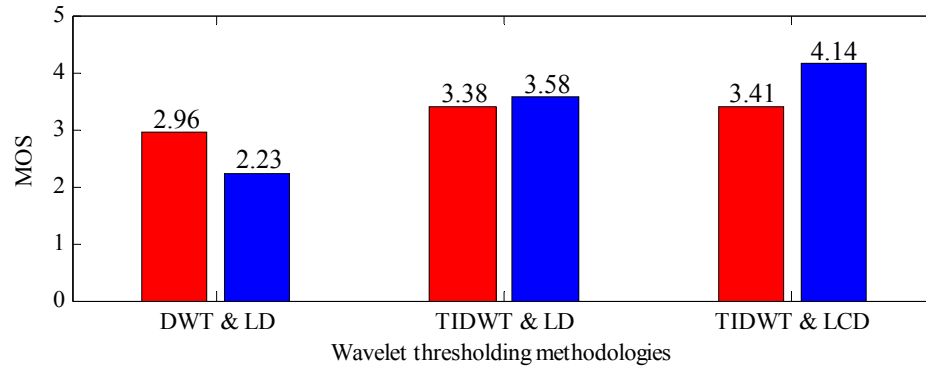
**Figure 4.7**    Mean opinion score (MOS) comparison for various wavelet thresholding methodologies with soft thresholding (red) and hard thresholding (blue): discrete wavelet transform (DWT) and level-dependent (LD) threshold, translation-invariant discrete wavelet transform (TIDWT) and LD threshold, and TIDWT and level-correlation-dependent (LCD) threshold.

perception than SNR value. Limited listeners for the MOS evaluation may also be a contributing factor. Nevertheless, it is evident from these results that the proposed LCD threshold with hard-thresholding rule under TIDWT setting performs best among the tested wavelet thresholding schemes for snore signal enhancement, both in SNR and MOS measures.

### 4.3.3    Snore Activity Detection

The robustness of the SA detector was comparatively evaluated with the segmentation subsystem developed in [147] for snoring episodes. In the subsystem, short-time energies and zero-crossing rates of signal frames of length 100 ms with 50% overlap were computed. Activity frames were marked when the energy and zero-crossing rate values of those frames were simultaneously above the recommended energy threshold and zero-crossing rate threshold. Boundaries of episodes were determined by continuities of activity frames, and episodes separated below certain duration were
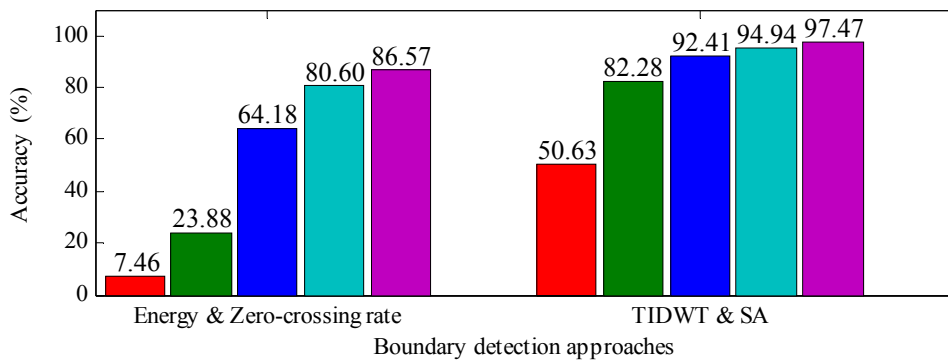
**Figure 4.8**     Accuracy comparisons between conventional (energy and zero-crossing rate) and wavelet-based (translation-invariant discrete wavelet transform (TIDWT) and snore activity (SA) detector) approaches in snore segment boundary detection at several tolerance degrees: ±25 ms (red), ±50 ms (green), ±75 ms (blue), ±100 ms (cyan), and ±120 ms (magenta).

merged. To access the effectiveness in identifying the snore segment boundaries, we compared the detected endpoints to the hand-labeled ones, with tolerance of ±25 ms, ±50 ms, ±75 ms, ±100 ms, and ±125 ms. For example, a snore segment was deemed 'correct' if the time difference between the detected endpoints and the manually labeled ones were within ±25 ms.

Figure 4.8 compares the performance of the conventional energy and zero-crossing rate approach adopted in [147] and the proposed SA detector in TIDWT domain. Within the tolerance of ±25 ms, 50.63% of the snore segments are correctly detected by the wavelet-based approach, whereas 7.46% by the conventional approach. On the other extreme, where the tolerance degree is ±125 ms, the wavelet-based approach yields an accuracy of 97.47%, while the conventional approach yields only 86.57%. A close inspection of Figure 4.8 shows that the wavelet-based SA detection is at least 10% more effective the conventional approach. This result lends support to our earlier hypothesis that good localization property in both the time and frequency domain of TIDWT may effectively aid SA detection.

**Table 4.3**     Diagnostic performance of spectral peak frequency (PF) using the fast Fourier transform of apneic and benign snores before and after preprocessing.

| Condition | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
|---|---|---|---|---|---|---|
| Before preprocessing | 232 | 0.6485 | 0.0424 | 45.3 | 86.0 | 0.0005 |
| After preprocessing | 240 | 0.8080 | 0.0311 | 76.7 | 82.0 | < 0.0001 |

Thre refers to threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

### 4.3.4    Clinical Efficacy

To examine the clinical usefulness of the proposed preprocessing system, we ran a simple trial comparing the diagnostic results from frequency analysis of the existing 400 inspiratory snores before and after preprocessing. The diagnostic performance was statistically assessed through the receiver operating characteristic (ROC) curve analysis [155-157] programmed in MATLAB$^{TM}$, and the results were confirmed by means of a statistical software package (MedCalc$^{TM}$, version 9.3.6.0, MedCalc Software). An optimal threshold value was determined based on maximum sum of sensitivity and specificity, along with an area under ROC curve (AUC), standard error of AUC (SE), sensitivity, specificity, and statistical significance (p-value). The p-value was calculated from a two-tailed z-test [139,140,157], with a null hypothesis for AUC = 0.5. The lower the p-value (< 0.05), the greater is the AUC difference from 0.5, and therefore, the better is the ability to differentiate between apneic and benign patients.

Following the analytical procedures in the study by Michael *el al.* [94], power spectrum of each snore before (as in [94]) and after preprocessing was generated from the FFT [113,114] with a Hanning window of 16384 samples (≈ 372 ms). The FFT is primarily formulated as

$$X_n = \sum_{k=1}^{N} x_k \exp\left(\frac{-j2\pi(k-1)(n-1)}{N}\right), \qquad \forall n \in [1\,;N] \tag{4.21}$$
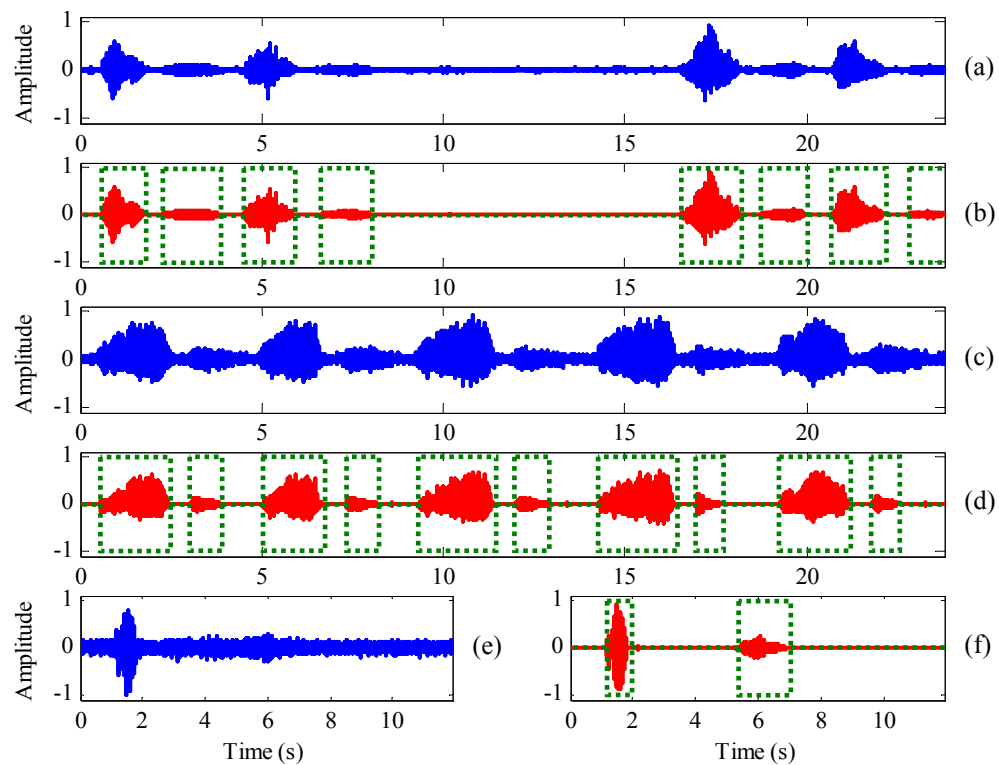
**Figure 4.9**    Snore signals of an apneic patient (a) before and (b) after preprocessing; a benign patient (c) before and (d) after preprocessing, and sleep sounds due to somniloquy and body movements (e) before and (f) after preprocessing. Segment boundaries are indicated in green.

where $\mathbf{x} = \{x_k\}_{k=1}^{N}$ is a snore signal with power-of-two length $N$, and $j = \sqrt{-1}$. Spectral PF was subsequently extracted to classify AS from BS. Diagnostic results for snores analyzed before and after preprocessing are listed in Table 4.3. While both produce evidences that PF of snore signals carry valuable information for OSA detection (p-value $\leq$ 0.0005), preprocessed snores deliver a higher diagnostic accuracy (AUC = 0.8080, SE = 0.0311, sensitivity = 76.7%, specificity = 82.0%, and p-value < 0.0001).

Apart from removing unwanted background noise and identifying snore segment boundaries, the proposed system can also helpfully mark out the endpoints of other sleep sounds due to somniloquy and body movements, as displayed in Figure 4.9, without influencing the above diagnostic outcomes.

## 4.4    Summary

This chapter proposes an advanced preprocessing system for snore signal enhancement and SA detection in a sleep laboratory environment. The proposed system concurrently performs two important tasks in a TIDWT domain: (a) wavelet-based denoising via a LCD threshold that tailors for colored noise embedded in snore signals; and (b) wavelet-based endpoint identification via a SA detector. Enhancement results demonstrate that the LCD threshold outperforms the LD threshold, both in terms of SNR and MOS, with an SNR increment between 3.02 dB and 38.22 dB over noisy snore signals. The robust statistical quality of the LCD threshold can be verified theoretically. In addition, results for detection of snore segment boundaries suggest that the proposed SA detector in TIDWT domain performs better than the conventional short-time energy and zero-crossing rate approaches in time-domain, with an accuracy of 50.63–97.47%, depending on the degree of tolerance. The proposed wavelet-driven system elegantly combines both enhancement and detection of snore signals in a TIDWT domain, resulting in lower computation costs and higher efficiency; this is the first reported in the literature. Diagnostic results also emphasize the clinical efficacy of the proposed preprocessing system, with higher accuracy for discriminating between AS and BS.

Upon preprocessing, snore signals can be readily subjected to further analysis and classification. The next chapter of this thesis presents a parametric approach for analyzing snore signals and to classify snores produced by apneic and benign patients.

Chapter 5

# Parametric Analysis and Classification of Snores

As mentioned in Section 1.1.2, snoring is generated by the vibrations of soft tissues and/or turbulence of airflow at constrictions in the UA. Similar to the vocal tract in speech production [152,158-161], the UA acts as a variable acoustic filter in the generation of snoring sounds: attenuating the transfer of sound energy at certain frequencies, while allowing maximal energy through at the resonant frequencies, also known as formant frequencies. Depending on the shape and physical dimensions of the UA, different snore sounds with diverse formant frequencies may be produced.

This chapter introduces a parametric approach to extract vital information about the UA anatomical structures through the estimation of formant signatures in snore signals. We hypothesize that formants of AS and BS may occupy different frequency ranges since patients with and without OSA have distinct differences in the UA anatomy [29-31,56,57]. Two popular techniques based on the parametric approach, namely linear prediction (LP) and discrete all-pole (DAP) modeling, are described, together with the algorithm for identifying formants in their respective spectrum. Experimental conditions, such as patient dataset and parameter selection procedures, are also highlighted. Subsequently, we presented and discussed the diagnostic accuracy of this approach for same- and both-gender patient groups through ROC curves and notched box plots. Besides that, linear and nonlinear regression models, in

---

The contents of this chapter are partly based on the author's publications [A2,A5,A14,A15,A16].
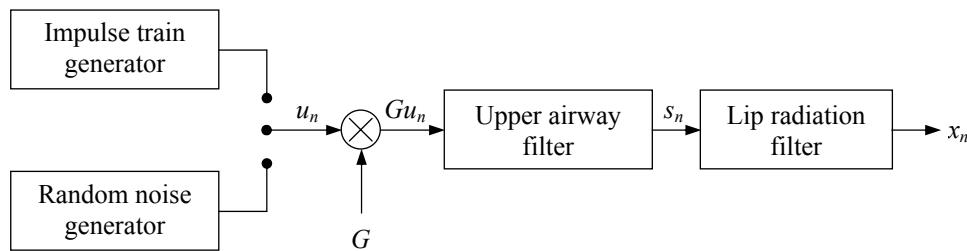
**Figure 5.1**   Source-filter model of snore production. $u_n$ and $G$ denote excitation source and gain, respectively. $H(z)$ denotes the transfer function of the upper airway filter, while $s_n$ and $x_n$ denote snore before and after lip radiation, respectively.

terms of linear, quadratic, logarithmic, exponential, and power functions, are defined to illustrate the correlation between AHI and the proposed diagnostic markers.

## 5.1    Source-Filter Theory for Snore Production

The source-filter theory [158,159], proposed in 1960, is a basic principle for interpreting speech phenomena and to develop speech technology applications [152,160,161] . The source-filter model works on an assumption that the sound excitation source and the filter are linearly separable [152,158-161]. In other words, the model obeys the superposition principle, meaning that the response of the system to a sum of inputs is the sum of the responses to each individual input. Owing to the fact that the biophysics of speech production and the generation of snoring sounds share many similarities (e.g., both speech and snore acoustic waves transmit through the same upper respiratory tract, and both have their spectral characteristics modulated by the respiratory tract anatomy [99,162]), it is reasonable to presume that the source-filter model may also apply to the generation of snoring sounds.

Figure 5.1 presents a source-filter model of snore production. As can be seen, the model comprises of an excitation source $u_n$, which can be excited by an impulse train (soft tissue vibrations) and/or a random sequence (noise-like turbulent airflow); a gain

Upper airway autoregressive filter

Upper airway
autoregressive filter

$$H(z) = \frac{1}{1 + \sum\limits_{k=1}^{p} a_k z^{-1}}$$
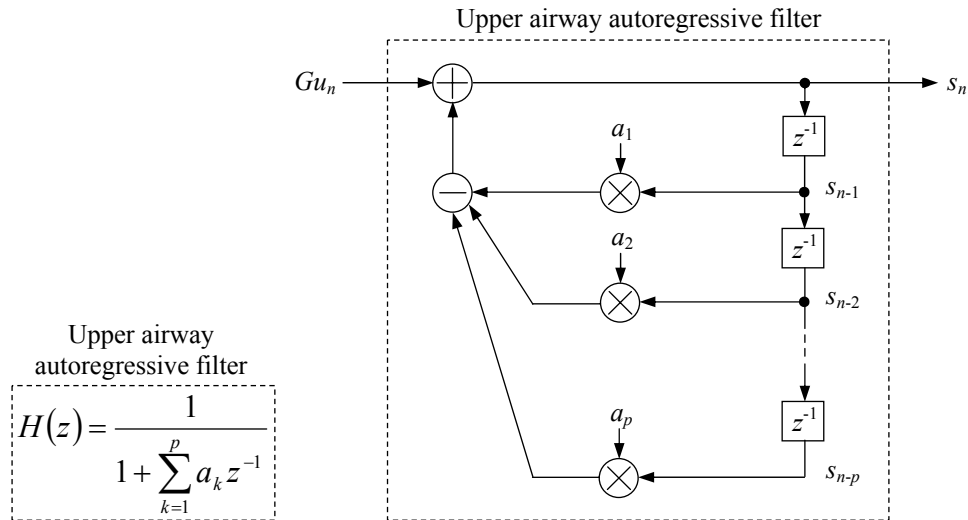
**Figure 5.2**   Linear prediction model for the upper airway. $H(z)$, $p$, and $a_k$ denote model transfer function, order, and coefficients, respectively.

$G$ that provides flexibility in the acoustic output level; a time-varying acoustic filter $H(z)$ for the UA; and a 1st-order differentiation filter for lip radiation [152,158-161]. In this representation, a snore $x_n$ is generated by passing a scaled excitation source through the variable UA filter and the lip radiation filter. Because the UA filter changes with time to simulate the effect of changes in the UA shape, it is the main ingredient in producing various snore sounds.

## 5.2    Linear Prediction

LP is a powerful parametric technique for estimating coefficients of a linear time-varying autoregressive filter [163-166], grounded in the source-filter theory. It is widely used in speech analysis, where coefficients are computed to define a vocal tract filter describing the speech signals [152,158-161,163-166] .

Figure 5.2 depicts a LP model for the UA. Consider a continuous-time signal $s(t)$ = $s(n\Delta t) = s_n$, where $n$ is an integer variable referring to the sample number and $\Delta t$ is the sampling period, LP works on the premise that adjacent signal samples are highly

correlated, allowing the signal $s_n$ to be a linear combination of past outputs and scaled input [163-166], i.e.,

$$s_n = -\sum_{k=1}^{p} a_k s_{n-k} + G u_n \tag{5.1}$$

where $p$ is the order of the LP model, whose transfer function is

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{5.2}$$

and coefficients are $\{a_k\}_{k=1}^{p}$. Since the input is usually unknown, the signal $s_n$ can only be approximately predicted, as follows:

$$\hat{s}_n = -\sum_{k=1}^{p} a_k s_{n-k} \ . \tag{5.3}$$

Thus, the prediction error is given by

$$e_n = s_n - \hat{s}_n = s_n + \sum_{k=1}^{p} a_k s_{n-k} \tag{5.4}$$

which implies that the smaller the error, the closer is the approximation, and hence, the better is the LP model representing the UA.

In accordance with the least-mean-square algorithm, commonly applied in adaptive filters to determine the filter coefficients [114,167], the LP coefficients $\{a_k\}_{k=1}^{p}$ can be found by minimizing the total squared error with respect to each of the coefficients. Explicitly,

$$\frac{\partial E}{\partial a_i} = 0, \qquad 1 \le i \le p \tag{5.5}$$

where

$$E = \sum_{n=-\infty}^{\infty} e_n^2 = \sum_{n=-\infty}^{\infty} \left( s_n + \sum_{k=1}^{p} a_k s_{n-k} \right)^2 \tag{5.6}$$

is the total squared error over an infinite duration.

In practice, we are interested in a finite-length signal. This is frequently achieved by multiplying the signal $s_n$ by a window function $w_n$ to obtain a windowed signal

$$s_n' = \begin{cases} s_n w_n & , \quad 0 \le n \le N-1 \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{5.7}$$

By constraining the analysis interval to the range $[0, N\text{-}1]$ and assuming the values are zero outside the range, the total squared error will be non-zero over the range $[0, N\text{-}1+p]$ and can be expressed as

$$E' = \sum_{n=0}^{N-1+p} \left( s_n' + \sum_{k=1}^{p} a_k s_{n-k}' \right)^2 . \tag{5.8}$$

Applying Equation (5.5) to Equation (5.8),

$$-\sum_{n=0}^{N-1+p} s_{n-i}' s_n' = \sum_{k=1}^{p} a_k \sum_{n=0}^{N-1+p} s_{n-i}' s_{n-k}' . \tag{5.9}$$

It can be recognized that the term $\sum_n s_{n-i}' s_{n-k}'$ is the covariance of $s_n'$,

$$\varphi(i,k) = \sum_{n=0}^{N-1+p} s_{n-i}' s_{n-k}' = \sum_{n=0}^{N-1-(i-k)} s_n' s_{n+i-k}' , \qquad 1 \le i \le p, \quad 0 \le k \le p . \tag{5.10}$$

Since the right-hand side of Equation (5.10) is identical to the short-time autocorrelation function evaluated for $(i-k)$, $\varphi(i,k) = R(i-k) = R(|i-k|)$ is an even function. Under these conditions, Equation (5.9) can be written as the Yule-Walker equations [168,169]

$$-R(i) = \sum_{k=1}^{p} a_k R(|i-k|), \qquad 1 \le i \le p \tag{5.11}$$

and in matrix notation as

$$
\begin{bmatrix}
R_0 & R_1 & \dots & R_{p-1} \\
R_1 & R_0 & \dots & R_{p-2} \\
\vdots & \vdots & \ddots & \vdots \\
R_{p-1} & R_{p-2} & \dots & R_0
\end{bmatrix}
\begin{bmatrix}
a_1 \\ a_2 \\ \vdots \\ a_p
\end{bmatrix}
=
\begin{bmatrix}
-R_1 \\ -R_2 \\ \vdots \\ -R_p
\end{bmatrix}
\tag{5.12}
$$

This formulation corresponds to an autocorrelation method, which is preferred over the covariance method for estimating the LP coefficients because it is computationally more efficient with $O(p^2)$ operations and guarantees a stable filter. The autocorrelation method yields a $p \times p$ symmetric Toeplitz matrix [170] that facilitates the applicability of Levinson-Durbin's recursive procedures [171] to solve the LP coefficients without direct matrix inversion, as specified below.

$$
E_0 = R(0) \tag{5.13a}
$$

$$
k_i = -\left[ R(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right] \Big/ E_{i-1} \tag{5.13b}
$$

$$
a_i^{(i)} = k_i \tag{5.13c}
$$

$$
a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}, \qquad 1 \le j \le i-1 \tag{5.13d}
$$

$$
E_i = \left(1 - k_i^2\right) E_{i-1}. \tag{5.13e}
$$

Equations (5.13b)–(5.13e) are recursively solved for $i = 1, 2,\dots, p$, and the coefficients of the $p$th-order LP model are

$$
a_j = a_j^{(p)}, \qquad 1 \le j \le p. \tag{5.14}
$$

In addition, the minimum total squared error $E_p$ is related to the gain $G$ as the total energy in the input

$$
G^2 = E_p = R(0) + \sum_{k=1}^{p} a_k R(k). \tag{5.15}
$$

Even though the estimation of autoregressive filter using LP has extensively employed in various speech processing applications (e.g., speech modeling and

recognition), the LP technique has its limitations: it inadequately models the vocal tract filters for high-pitched sounds and voices [165,172,173]. Consequently, the LP envelope fails to match the signal spectrum, and its peaks (formants) tend to drift towards the pitch harmonics, causing substantial inaccuracies in formant frequencies. This is attributed to (a) the aliasing occurring in the autocorrelation method, which grows with increasing fundamental frequency of the excitation source; and (b) the error cancellation property inherent in the LP error criterion [172,173].

As highlighted earlier, the error criterion for LP is based on the least squares distance between the actual and the predicted values. In the frequency domain, the LP error criterion for discrete spectra is equivalent to minimizing [172,173]

$$E_{\mathrm{LP}} = \frac{1}{N} \sum_{n=1}^{N} \frac{P(w_n)}{\hat{P}_{\mathrm{LP}}(w_n)} \tag{5.16}$$

where

$$\hat{P}_{\mathrm{LP}}(w) = \left| \hat{H}_{\mathrm{LP}}(w) \right|^2 = \frac{1}{\left| \sum_{k=0}^{p} a_k \exp(-jwk) \right|^2} . \tag{5.17}$$

is the spectrum of LP envelope, whereas $P(w)$ is the spectrum of the signal. The term $N$ is the number of a set of discrete frequencies $w_n$ with the range $-\pi \le w_n \le \pi$, and $j = \sqrt{-1}$. The gain $G$ is included in the coefficients of the denominator; therefore, $a_0$ is not confined to 1. Minimization of $E_{\mathrm{LP}}$ with respect to the LP coefficients $\{a_k\}_{k=1}^{p}$ yields the Yule-Walker equations

$$\sum_{k=0}^{p} a_k R(i-k) = 0, \qquad 1 \le i \le p \tag{5.18}$$

and
$$\sum_{k=0}^{p} a_k R(k) = \frac{1}{a_0} \tag{5.19}$$

where

$$R(i) = \frac{1}{N} \sum_{n=1}^{N} P(w_n) \exp(jw_n i) \qquad (5.20)$$

is the autocorrelation of the discrete signal spectrum $P(w_n)$. The minimization process

sets the autocorrelation of the continuous LP envelope

$$\hat{R}_{LP}(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{P}_{LP}(w) \exp(jwi)\, dw \qquad (5.21)$$

to that of the discrete signal spectrum, i.e.,

$$\hat{R}_{LP}(i) = R(i), \qquad 0 \le i \le p. \qquad (5.22)$$

Substituting the spectrum of the signal

$$P(w) = \sum_{l=-\infty}^{\infty} R_s(l) \exp(-jwl) \qquad (5.23)$$

where

$$R_s(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(w) \exp(jwi)\, dw \qquad (5.24)$$

is the autocorrelation of the signal spectrum into Equation (5.20) yields the

relationship between $R$ and $R_s$:

$$R(i) = \frac{1}{N} \sum_{n=1}^{N} \sum_{l=-\infty}^{\infty} R_s(l) \exp(-jw_n(l-i)), \qquad \forall i. \qquad (5.25)$$

For a periodic excitation source, $w_n = 2\pi(n-1)/N$, where $N$ denotes the period of

excitation, Equation (5.25) reduces to

$$R(i) = \sum_{l=-\infty}^{\infty} R_s(i - lN), \qquad \forall i \qquad (5.26)$$

which indicates the presence of aliasing in the autocorrelation domain whenever a

spectral envelope is sampled at a set of discrete frequencies. By further examining the

above equations, one can also deduce that

$$\hat{R}_{LP}(i) = R(i) = \sum_{l=-\infty}^{\infty} R_s(i - lN) \ne R_s(i), \qquad 0 \le i \le p. \qquad (5.27)$$

In essence, LP fails to compensate for the aliasing effect since the autocorrelation corresponding to the LP envelope will always equal an aliased version of the signal spectrum rather than the original signal spectral envelope. The autocorrelation aliasing becomes more severe when the pitch increases as there are fewer harmonics, thereby degrading the LP modeling accuracy for high-pitched sounds and voices. Another explanation for the inadequacy of LP envelope in matching the original envelope is the error cancellation property [172,173] possessed by the LP error criterion in Equation (5.16). The errors due to poor fitting when $P(w) > \hat{P}_{\text{LP}}(w)$ are cancelled by those errors when $\hat{P}_{\text{LP}}(w) > P(w)$, resulting LP to select an envelope other than the ideal one passing through all the spectral points.

## 5.3    Discrete All-Pole Modeling

The drawbacks of LP have led to the introduction of DAP modeling [173], with an objective to resolve the aliasing problem faced by LP, and hence, to correctly estimate the autoregressive filter. This objective is achieved by employing a discrete version of the Itakura-Saito error measure [174,175]

$$E_{\text{DAP}} = \frac{1}{N} \sum_{n=1}^{N} \left[ \frac{P(w_n)}{\hat{P}_{\text{DAP}}(w_n)} - \log_{\text{e}} \left( \frac{P(w_n)}{\hat{P}_{\text{DAP}}(w_n)} \right) - 1 \right] \qquad (5.28)$$

where $P(w_n)$ is the discrete signal spectrum defined at $N$ discrete frequencies $w_n$ with the range $-\pi \le w_n \le \pi$, and $\hat{P}_{\text{DAP}}(w_n)$ is the DAP model spectrum. Unlike $E_{\text{LP}}$ in Equation (5.16), $E_{\text{DAP}}$ has the prefect-fit property that is always non-negative and gives a minimum value (zero) only when $P(w_n) = \hat{P}_{\text{DAP}}(w_n)$. Moreover, this error measure does not forgo any of the advantages of LP (e.g., perfect modeling for low periodic and random excitation sources) [173].

To satisfy the condition for minimizing $E_{DAP}$, Equation (5.17) is rewritten as

$$\hat{P}_{DAP}(w) = \left| \hat{H}_{DAP}(w) \right|^2 = \frac{1}{\sum_{k=0}^{p} d_k \cos(wk)} \tag{5.29}$$

where

$$d_0 = \sum_{k=0}^{p} a_k^2 \tag{5.30}$$

and

$$d_i = 2 \sum_{k=0}^{p-i} a_k a_{k+i}, \qquad 1 \le i \le p. \tag{5.31}$$

Substituting Equation (5.29) into Equation (5.28), and then setting

$$\frac{\partial E_{DAP}}{\partial d_i} = 0, \qquad 1 \le i \le p \tag{5.32}$$

yields

$$\hat{R}_{DAP}(i) = R(i), \qquad 0 \le i \le p \tag{5.33}$$

which appears deceptively identical to Equation (5.22). The main different is that $\hat{R}_{LP}(i)$ is the autocorrelation of the continuous LP spectrum $\hat{P}_{LP}(w)$, whereas

$$\hat{R}_{DAP}(i) = \frac{1}{N} \sum_{n=1}^{N} \hat{P}_{DAP}(w_n) \exp(jw_n i) \tag{5.34}$$

is the autocorrelation of the DAP spectrum sampled at the same discrete frequencies as the signal spectrum. Accordingly, DAP modeling can suitably match the aliased autocorrelation of the discrete signal spectrum to the autocorrelation of the DAP model aliased in the same way. Thus, it is superior to LP in analyzing periodic excitation source; however, at the price of increased computational costs and algorithmic complexity because there is no closed-form solution for the DAP coefficients and an iterative method [173] must be implemented, as detailed below.

DAP coefficients are obtained by using the definition in Equation (5.17) and setting

$$\frac{\partial E_{\text{DAP}}}{\partial a_i} = 0, \qquad 0 \le i \le p. \tag{5.35}$$

This produces a set of nonlinear equations

$$\sum_{k=0}^{p} a_k R(i-k) = \sum_{k=0}^{p} a_k \hat{R}_{\text{DAP}}(i-k), \qquad 0 \le i \le p. \tag{5.36}$$

For simplicity, a property of autoregressive filter, formulated as

$$\sum_{k=0}^{p} a_k \hat{R}(i-k) = \hat{h}(-i), \qquad \forall i \tag{5.37}$$

is applied to compute the DAP coefficients $\{a_k\}_{k=0}^{p}$. The term

$$\hat{h}(-i) = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{\exp(-jw_n i)}{\sum_{k=0}^{p} a_k \exp(-jw_n k)} \right), \qquad \forall i \tag{5.38}$$

is the time-reversed impulse response of the discrete frequency sampled autoregressive

model and a function of the coefficients, which yields

$$\sum_{k=0}^{p} a_k R(i-k) = \hat{h}(-i), \qquad 0 \le i \le p \tag{5.39}$$

when substituting Equation (5.37) into Equation (5.36). The DAP coefficients are
eventually solved by the following steps [173]:

(a)    perform peak picking on the signal spectrum to yield the locations $w_n$, the

      magnitudes $P(w_n)$, and the number $N$ of peaks;

(b)    calculate $R(i)$ based on Equation (5.20);

(c)    find an initial estimates of $\{a_k\}_{k=0}^{p}$ using ordinary linear prediction;

(d)    determine $\hat{h}(-i)$ using Equation (5.38);

(e)    solve Equation (5.39) for new $\{a_k\}_{k=0}^{p}$ estimates;

(f)    examine $E_{\text{DAP}}$ using Equation (5.17) and Equation (5.28);

(g)    return to step (d) if the reduction of $E_{DAP}$ from previous iterations is greater than some threshold value (e.g., 0.0001); else continue

(h)    normalize $\{a_k\}_{k=0}^p$ so that the energy in the residual spectrum

$$\frac{1}{N}\sum_{n=1}^{N}\frac{P(w_n)}{\hat{P}_{DAP}(w_n)}=1\,;\text{ and}$$

(g)    stop.

## 5.4    Formant Extraction

Formants are resonances in the UA. They manifest as energy peaks (maxima) at the formant frequencies in both LP and DAP spectra. Frequencies at which the formants occur rely upon the shape and size of the UA; therefore, it is important to appropriately identify formants, especially the first three formants, in order to better describe the UA configuration. Two typical algorithms for formant extraction are the root-finding and the peak-picking [152,160,176,177].

The root-finding algorithm approximates formant frequencies by means of solving the poles of the autoregressive filter modeling the UA. This is achieved by setting the denominator of the filter to zero and determining the roots of the $p$th-order polynomial. The denominator polynomial can be expressed as a product of basic terms $(1-z_iz^{-1})$, where $z_i = r_i\exp(j\theta_i)$ is a pole in the z-domain with frequency

$$F_i = \frac{f_S}{2\pi}\theta_i \tag{5.40}$$

and a 3 dB bandwidth

$$B_i = -\frac{f_S}{\pi}\log_e(r_i) \tag{5.41}$$

where $r_i$ is the magnitude of the pole, $\theta_i$ is the phase, and $f$s is the sampling frequency

[152,160,176,177]. The root-finding algorithm is an obvious approach towards formant extraction, yet it suffers from several shortcomings. Firstly, it is computational burdensome to obtain roots of the $p$th-order polynomial [176] despite the aids of numerical methods (e.g., Laguerre method and Muller method [178]). Secondly, it is difficult to tell whether an obtained root contributes to a true formant since the poles are susceptible to pole interaction [179], which often arises when poles are close together. Finally, selection of pole pairs can be tedious as it takes into an account the formant frequency location, bandwidth, and continuity [177].

The peak-picking algorithm, on the other hand, merely chooses the peaks in the LP or DAP spectrum as formants: the first peak as first formant (F1), the second as second formant (F2), and the third as third formant (F3), beginning with the lowest frequency, with an assumption that a pole is capable of manifesting itself as a peak in the spectrum. Although this assumption holds most of the time, errors can still occur when (a) two poles appear as a single peak when they are close to one another in frequency; and (b) a pole attributable to frequency shaping emerge as a spurious peak, which would be mistakenly identified as a formant [177]. Conversely, the algorithm is simple and requires less computation. A study has reported that, for up to 90% of the time, the first three peaks of the spectrum can be truly interpreted as the first three formants [180], thereby augmenting the reliability of peak-picking approach. Furthermore, the algorithm is tractable as spurious peaks are rare [177].

Due to these advantages of peak picking, we chose the peak-picking rather than the root-finding algorithm for extracting the first three formant frequencies (F1, F2, and F3) from both LP and DAP spectra. To accomplish this task, we developed a peak-picking program that successfully tracks the gradient of the spectral envelope and recognizes a peak when there is a change of gradient from positive to negative.

## 5.5     Experimental Conditions

### 5.5.1     Patient Dataset

Snoring sounds of the 40 patients (30 apneic and 10 benign) listed in Table 1.1 were simultaneously recorded with PSG through the snore signal acquisition system (see Chapter 3) in the sleep laboratory. The acquired snore signals were preprocessed via the wavelet-based preprocessing system (see Chapter 4) to enhance the signal quality and intelligibility, along with the detection of snore sounds onset.

The preprocessed snores were further analyzed by Cool Edit Pro$^{TM}$. For each patient, 40 inspiratory snores of an average root-mean-square power of -26.95 $\pm$ 6.34 dBFS were selected over 6.58 $\pm$ 0.96 h continuous recording. The first 30 snores were designated as training data for computing the formant frequencies (F1, F2, and F3) and to derive optimal threshold values using the ROC curve analysis [155-157] elaborated in Section 4.3.4. AUCs and their corresponding standard errors, together with p-values (two-tailed z-test [139,140,157]) were also calculated. A p-value < 0.05 was considered statistical significant.

In addition, the training dataset was evaluated by means of notched box plot [139,140,181,182], a useful graphical method for exploratory data analysis. In a notched box plot, the central box specifies the values from 25 to 75 percentile. The notched area is the 95% confidence limits for the median indicated by a middle line, while the whiskers show the rest of the data that are within 1.5 times the interquartile range from the lower or upper quartile values. Outlines are also highlighted by crosses.

The remaining 10 snores were designated as test data to assess the sensitivity and specificity of the derived threshold values. In total, the dataset contains 900 AS and 300 BS for training, and 300 AS and 100 BS for testing. Naturalness of these snores was also subjectively validated by the group of polysomnographic technicians

and signal processing specialists [25,183].

To increase computational efficiency in the subsequent analysis, the snores were first low-pass filtered at a cutoff frequency of 5000 Hz via an 8th-order Butterworth anti-aliasing filter [114] with a zero-phase filtering function [184] that eliminates the nonlinear phase distortion inherent in the infinite impulse response Butterworth filter, and then downsampled by a factor of 4, yielding a new sampling frequency $f_s = 11025$ Hz. This new sampling frequency fulfills the Nyquist criterion [113,114] since the maximum frequency of interest for snore signals is about 5000 Hz [99,105].

## 5.5.2    Optimal Parameter Selection

Prior to LP or DAP analysis, several issues concerning preemphasis and windowing, as well as the autoregressive filter order, are to be addressed [164]. Figure 5.3a illustrates the FFT spectrum of a frame size snore signal and its corresponding LP and DAP spectral envelopes. The signal exhibits a high-frequency roll-off. This phenomenon, which is commonly seen in speech signals, is likely attributed to the combined effect of -6 dB/octave of the excitation source (-12 dB/octave) and the lip radiation (+6 dB/octave). Hence, it is desirable to flatten the spectrum before LP or DAP analysis. Compensation for the high-frequency loss is often achieved by applying a preemphasis filter

$$g(z) = 1 - bz^{-1}, \qquad 0.9 \le b \le 1.0 \tag{5.42}$$

with a familiar choice of $b = 15/16 = 0.9375$ [164,185-187]. Accordingly, if $x_n$ is the input denoised snore signal, the preemphasized signal is

$$s_n = x_n - 0.9375x_{n-1} \tag{5.43}$$

where $n$ is an integer variable. The plots of the same frame size signal after preemphasis are illustrated in Figure 5.3b. As can be seen, the preemphasized signal
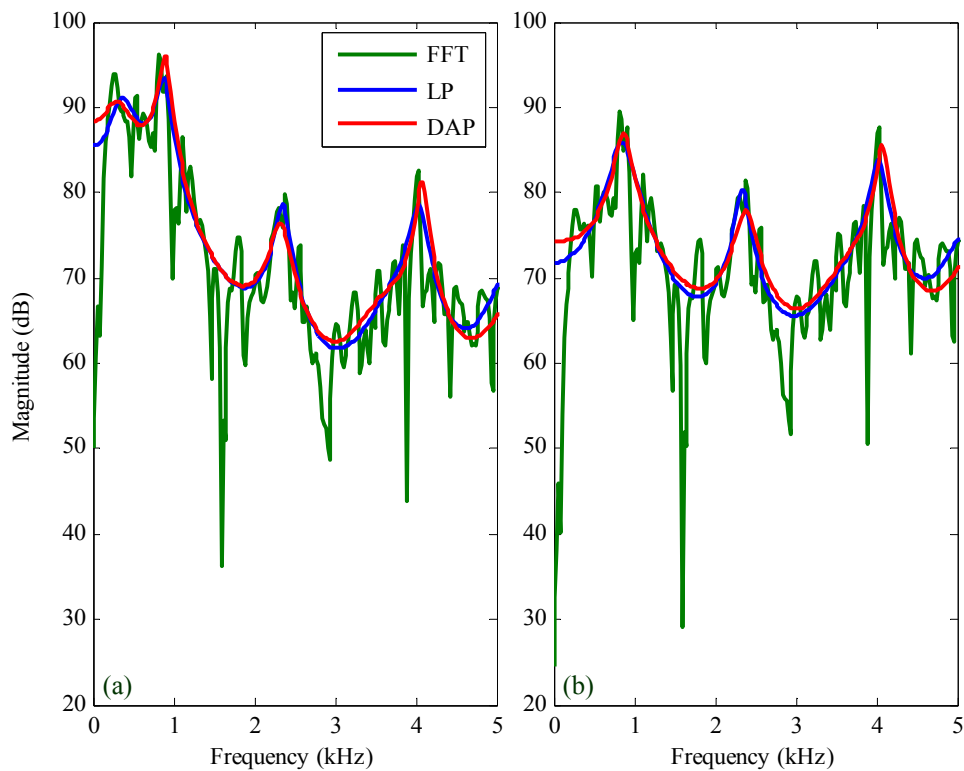
**Figure 5.3**     Spectra from fast Fourier transform (FFT), linear prediction (LP), and discrete all-pole (DAP) modeling of a snore signal (a) before and (b) after preemphasis.

spectrum is equalized; magnitudes of high-frequency components are boosted to be on par with the low-frequency ones. Preemphasized signals are also less subjected to finite-precision effects [186,187], such as underflow and overflow.

Windowing is essential in the autocorrelation method of LP and DAP modeling. Application of windowing segments a signal into small finite-length frames that are quasi-stationary, and it avoids rapid changes of signal at the window edge, ensuring a smooth transition between the successive frames. We selected the Hanning window function given in Equation (3.6) because it is well-liked, simple to execute, and has an acceptable sidelobe behavior [113-115]. Since the usual fundamental frequency of snores can be as low as 50 Hz ($\approx$ 20 ms), we decided on a frame size of 256 samples ($\approx$ 23 ms) with an overlap of 75% between consecutive frames. These selections are also highly recommended for speech analysis, where the frame size normally spans in

the range of 20–30 ms with 50% or 75% overlap [164,188,189]. Moreover, the selections are consistent with those in Chapters 3 and 4.

Besides the abovementioned issues, a major consideration prior to LP or DAP analysis is the order of the autoregressive filter representing the UA. In other words, the order of LP or DAP model defines the UA resonances, which is not possibly known due to the complexity of the UA configuration. Nevertheless, the model order $p$ can be reasonably estimated as follows [166]:

$$p \approx 2\left(N_{\mathrm{F}}\right) + \left(2 \text{ to } 3\right) \tag{5.44}$$

where
$$N_{\mathrm{F}} = \frac{f_{\mathrm{S}}}{2\left(F_{n+1} - F_n\right)} \tag{5.45}$$

denotes the number of formants in the spectrum. The extra 2 to 3 poles account for spectral tilt and provide spectral balance. The term $f$s is the sampling frequency in hertz, whereas

$$F_{n+1} - F_n \approx \frac{c}{2l} \tag{5.46}$$

is the approximate spacing between neighboring formant frequencies

$$F_n = \frac{(2n-1)c}{4l}, \qquad n \in \mathrm{N} \tag{5.47}$$

where $c$ denotes the speed of sound and $l$ denotes the UA length. Since $f$s = 11025 Hz, $c \approx 35400$ cm/s for moist air at 37ºC, the possible model orders for (a) UA length of 18.7 cm (from larynx to lips in adults [99]) are 14 to 15; (b) UA length of 19.6 cm (from mid-trachea portion to upper incisor in males [190]) are 14 to 15; and (c) UA length of 17.7 cm (from mid-trachea portion to upper incisor in females [190]) are 13 to 14. Based on these outcomes, we configured the UA by LP or DAP model with $p$ = 14, which is also a favorable opinion for the autoregressive filter describing the speech signals or breath sounds [160,166,191].
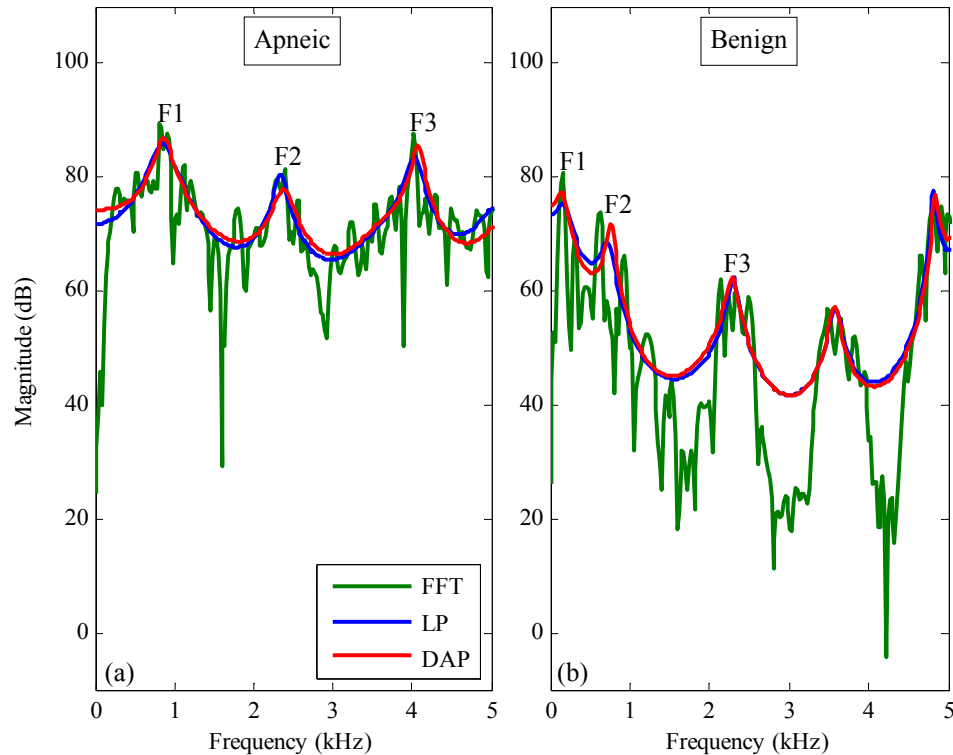
**Figure 5.4**   Spectra from fast Fourier transform (FFT), linear prediction (LP), and discrete all-pole (DAP) modeling of a typical (a) apneic snore and (b) benign snore.


## 5.6    Diagnostic Results and Discussion

Figure 5.4 exemplifies the LP and DAP spectral envelopes of a typical apneic snore and benign snore. It is obvious from the plots that formants of the apneic snore (F1 = 840–861 Hz, F2 = 2347–2390 Hz, F3 = 4027–4070 Hz) have higher frequencies than that of the benign snore (F1 = 150–151 Hz, F2 = 711–754 Hz, F3 = 2283–2304 Hz).

Based on the existing training dataset, mean and standard deviation of the formant frequencies (F1, F2, and F3) extracted from LP and DAP spectra are summarized in Table 5.1 and Table 5.2, respectively. Results consistently demonstrate quantitative differences in the formant frequencies of AS and BS for all three patient groups (males, females, and combined), with AS occupying higher frequency ranges than BS. These discrepancies can be explained by the differences in UA anatomical

**Table 5.1**  Diagnostic performance of formant frequencies (F1, F2, and F3) computed using linear prediction of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

| | | | | F1 | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
| M:A | 52.8 ± 24.7 | 750 ± 204 (239–1425) | 545 | 0.8977 | 0.0088 | 82.5 | 91.7 | < 0.0001 |
| M:B | 5.3 ± 3.5 | 425 ± 157 (155–852) | | | | | | |
| F:A | 23.6 ± 15.1 | 623 ± 150 (299–1120) | 298 | 0.8969 | 0.0088 | 100 | 95.0 | < 0.0001 |
| F:B | 3.4 ± 3.2 | 263 ± 187 (133–854) | | | | | | |
| C:A | 46.9 ± 25.7 | 724 ± 201 (239–1425) | 470 | 0.8992 | 0.0087 | 88.3 | 82.0 | < 0.0001 |
| C:B | 4.6 ± 3.4 | 360 ± 187 (133–854) | | | | | | |

| | | | | F2 | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
| M:A | 52.8 ± 24.7 | 1842 ± 238 (1062–2629) | 1808 | 0.6860 | 0.0164 | 58.3 | 55.0 | < 0.0001 |
| M:B | 5.3 ± 3.5 | 1670 ± 271 (837–2268) | | | | | | |
| F:A | 23.6 ± 15.1 | 1676 ± 303 (1080–2288) | 1380 | 0.5894 | 0.0183 | 83.3 | 55.0 | < 0.0001 |
| F:B | 3.4 ± 3.2 | 1564 ± 294 (898–2366) | | | | | | |
| C:A | 46.9 ± 25.7 | 1809 ± 261 (1062–2629) | 1740 | 0.6820 | 0.0165 | 65.0 | 53.0 | < 0.0001 |
| C:B | 4.6 ± 3.4 | 1627 ± 285 (837–2366) | | | | | | |

| | | | | F3 | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
| M:A | 52.8 ± 24.7 | 3008 ± 266 (2036–3880) | 2705 | 0.6117 | 0.0179 | 88.8 | 31.7 | < 0.0001 |
| M:B | 5.3 ± 3.5 | 2858 ± 405 (1319–3751) | | | | | | |
| F:A | 23.6 ± 15.1 | 2739 ± 282 (2272–3492) | 2439 | 0.4160 | 0.0194 | 96.7 | 37.5 | 2.0000 |
| F:B | 3.4 ± 3.2 | 2813 ± 297 (2210–3987) | | | | | | |
| C:A | 46.9 ± 25.7 | 2955 ± 290 (1062–3880) | 2798 | 0.5964 | 0.0182 | 77.3 | 49.0 | < 0.0001 |
| C:B | 4.6 ± 3.4 | 2840 ± 366 (837–3987) | | | | | | |

Values are presented as mean ± standard deviation and range within brackets. AHI refers to apnea-hypopnea index in events/h; Thre, threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

**Table 5.2**   Diagnostic performance of formant frequencies (F1, F2, and F3) computed using discrete all-pole modeling of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
|------|-----|-----------|-----------|-----|-----|---------|----------|---------|
| | | **F1** | | | | | | |
| M:A | 52.8 ± 24.7 | 650 ± 103 (377–1002) | 541 | 0.8457 | 0.0112 | 83.8 | 85.0 | < 0.0001 |
| M:B | 5.3 ± 3.5 | 454 ± 154 (144–835) | | | | | | |
| F:A | 23.6 ± 15.1 | 620 ± 98 (291–1006) | 375 | 0.9231 | 0.0074 | 100 | 92.5 | < 0.0001 |
| F:B | 3.4 ± 3.2 | 253 ± 171 (106–816) | | | | | | |
| C:A | 46.9 ± 25.7 | 644 ± 103 (291–1006) | 493 | 0.8768 | 0.0098 | 89.0 | 82.0 | < 0.0001 |
| C:B | 4.6 ± 3.4 | 374 ± 189 (106–835) | | | | | | |
| | | **F2** | | | | | | |
| M:A | 52.8 ± 24.7 | 1706 ± 202 (1138–2285) | 1528 | 0.5280 | 0.0190 | 79.6 | 38.3 | 0.1419 |
| M:B | 5.3 ± 3.5 | 1677 ± 246 (817–2224) | | | | | | |
| F:A | 23.6 ± 15.1 | 1668 ± 273 (1142–2356) | 1380 | 0.6099 | 0.0180 | 86.7 | 47.5 | < 0.0001 |
| F:B | 3.4 ± 3.2 | 1538 ± 331 (703–2261) | | | | | | |
| C:A | 46.9 ± 25.7 | 1698 ± 218 (1138–2356) | 1405 | 0.5718 | 0.0185 | 91.7 | 32.0 | < 0.0001 |
| C:B | 4.6 ± 3.4 | 1621 ± 290 (703–2261) | | | | | | |
| | | **F3** | | | | | | |
| M:A | 52.8 ± 24.7 | 2797 ± 278 (2132–3512) | 2374 | 0.4543 | 0.0194 | 95.0 | 16.7 | 1.9813 |
| M:B | 5.3 ± 3.5 | 2841 ± 377 (1292–3777) | | | | | | |
| F:A | 23.6 ± 15.1 | 2668 ± 291 (2274–3581) | 2250 | 0.3813 | 0.0193 | 100 | 15.0 | 2.0000 |
| F:B | 3.4 ± 3.2 | 2777 ± 330 (2166–3886) | | | | | | |
| C:A | 46.9 ± 25.7 | 2771 ± 285 (1138–3581) | 2374 | 0.4587 | 0.0194 | 95.0 | 21.0 | 1.9665 |
| C:B | 4.6 ± 3.4 | 2815 ± 360 (703–3886) | | | | | | |

Values are presented as mean ± standard deviation and range within brackets. AHI refers to apnea-hypopnea index in events/h; Thre, threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

structures in patients with and without OSA. Patients with OSA usually have narrower and unstable UAs that collapse easily than those without OSA [29-31,56,57].

Apart from that, we observed that the formant frequencies for male patients are generally higher than female patients in their respective groupings. This is perhaps owing to (a) an anatomically less stable UA, which is more prone to OSA in males than females [192-196], or (b) simply that the sample size for female subjects (6 apneic and 4 benign) is too small to be conclusive.

For each patient group, the potential diagnostic value of formant frequencies was further examined through ROC curves displayed in Figure 5.5 to determine threshold values, along with their corresponding AUCs, SEs, and p-values, as presented in Table 5.1 and Table 5.2. To facilitate easy interpretation, the classification results are graphically depicted in Figures 5.6 and 5.7 via notched box plots. The plots for both males and females combined clearly indicate that the threshold of F1 (LP; DAP) = (470; 493) Hz can best classify AS from BS, with a resultant AUC (LP; DAP) = (0.8992; 0.8768), SE (LP; DAP) = (0.0087; 0.0098), and p-value (LP; DAP) = (< 0.0001; < 0.0001). This threshold value is also found to lie between the threshold for males, which is F1 (LP; DAP) = (545; 541) Hz, and for females, which is F1 (LP; DAP) = (298; 375) Hz.

Each of the derived threshold values was applied to the corresponding test dataset to assess its sensitivity and specificity. Table 5.1 and Table 5.2 reveal that, for both-gender patient group, the yielding sensitivity and specificity are (LP; DAP) = (88.3%; 89.0%) and (LP; DAP) = (82.0%; 82.0%), respectively. In all three groupings, the thresholds for F1 constantly achieve higher sensitivity (LP: 82.5–100%; DAP: 83.8–100%) and specificity (LP: 82.0–95.0%; DAP: 82.0–92.5%) than F2 and F3. Therefore, one can infer that F1 of snore signals contain rich information to detect
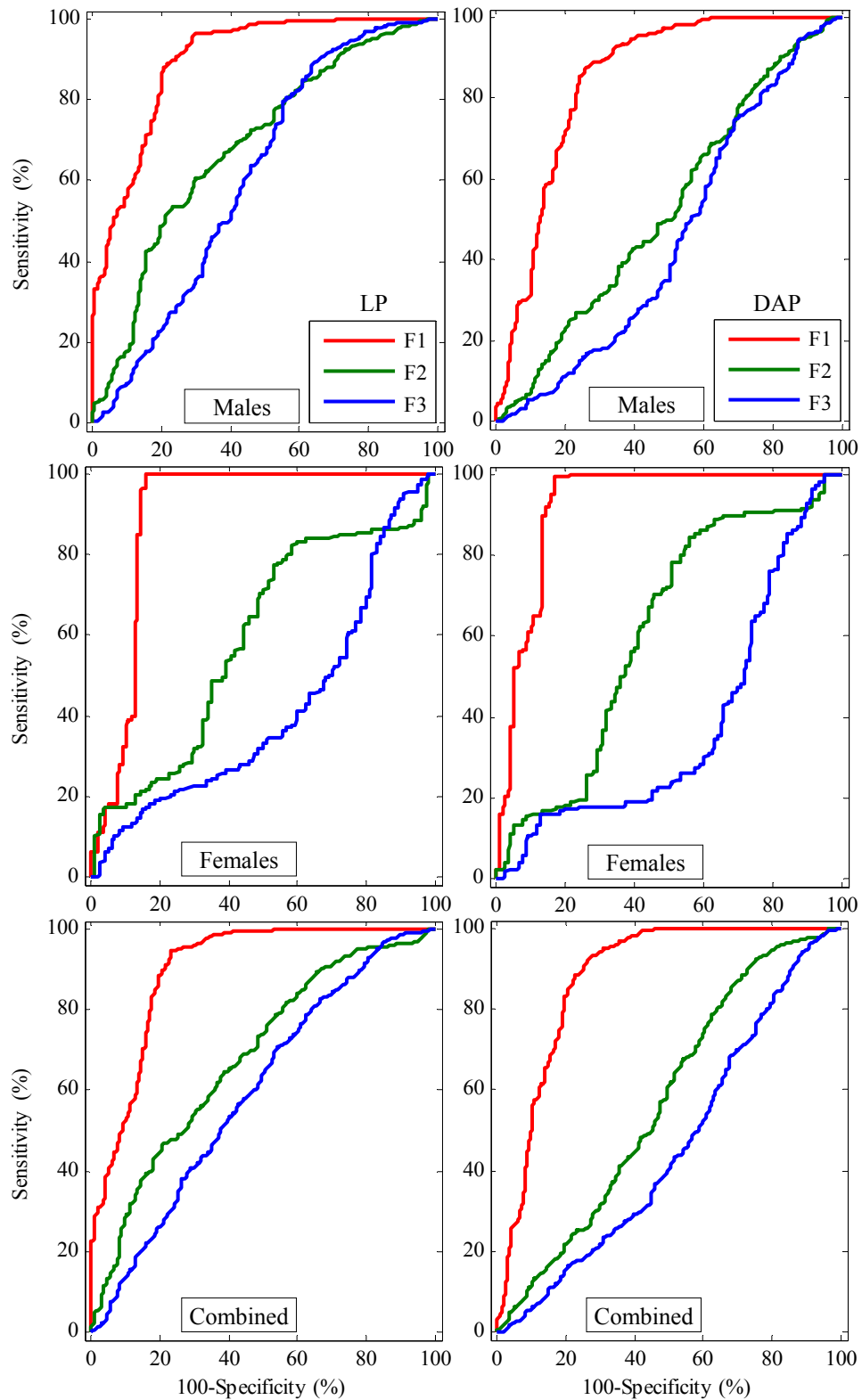
**Figure 5.5**    Receiver operating characteristic curves of formant frequencies (F1, F2, and F3) computed using linear prediction (left column) and discrete all-pole modeling (right column) of apneic and benign snores for males, females, and both males and females combined.
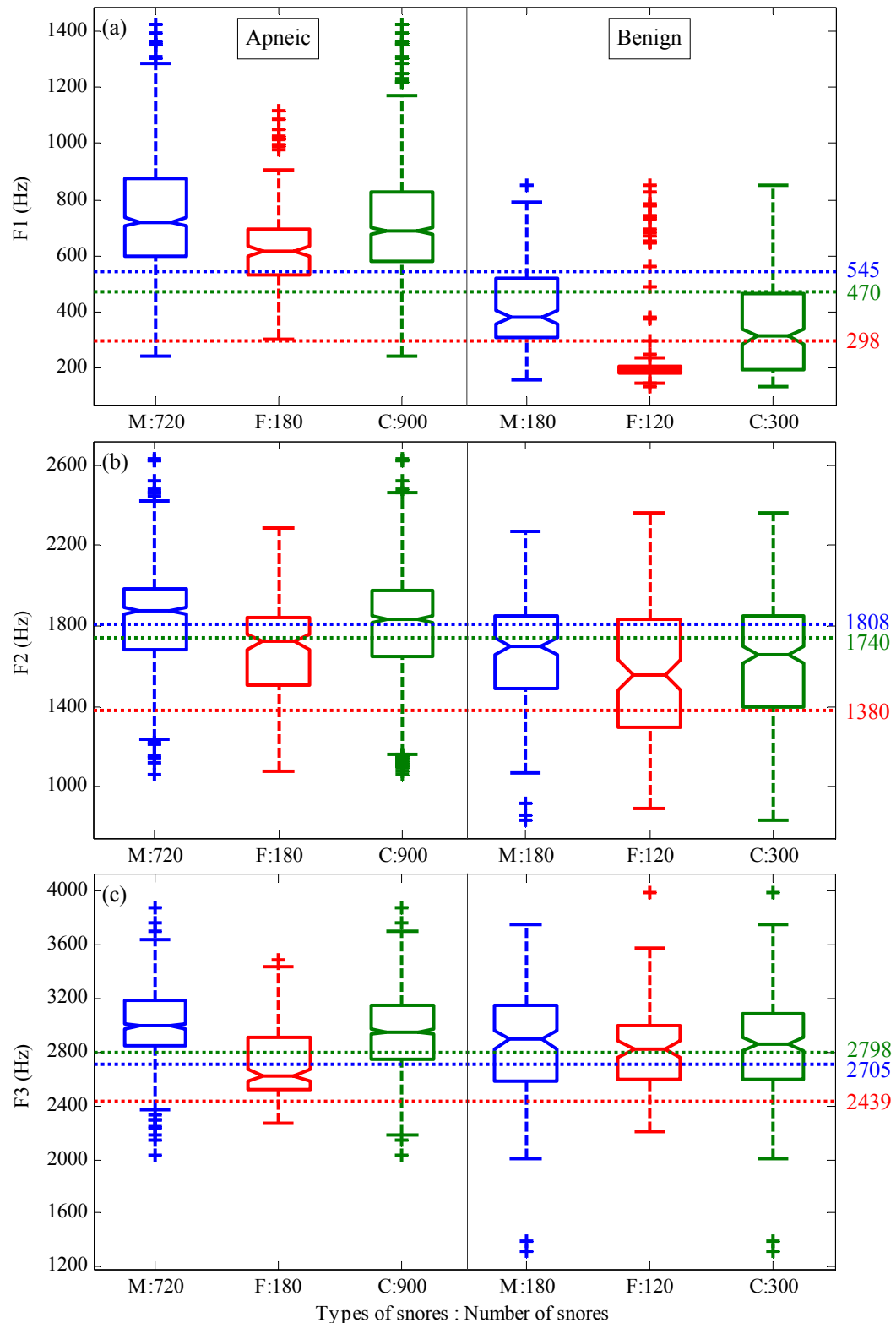
**Figure 5.6**  Notched box plots of formant frequencies: (a) F1, (b) F2, and (c) F3, computed using linear prediction of apneic and benign snores for males (M), females (F), and both males and females combined (C). Threshold values are marked on the right-hand side of the plots.
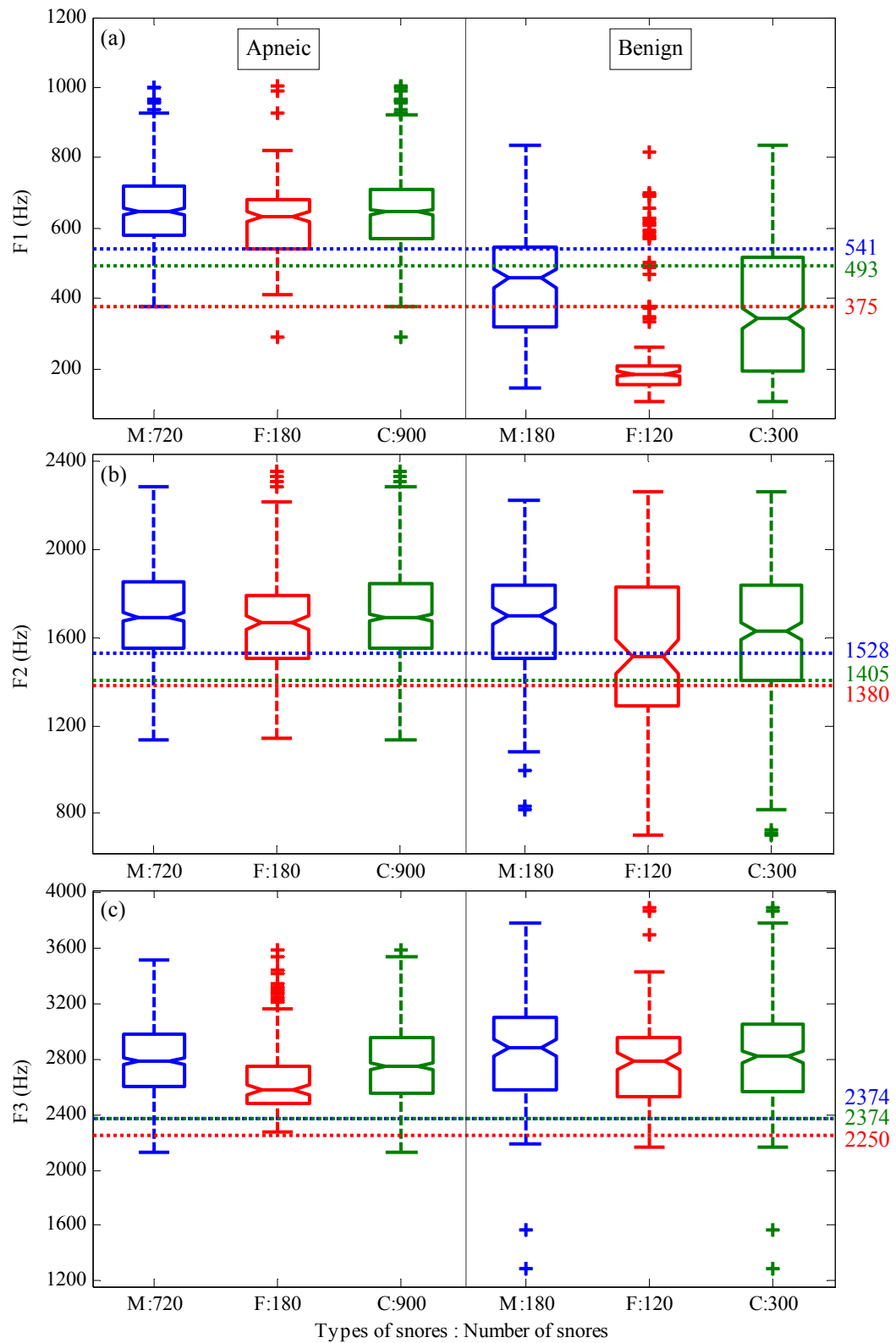
**Figure 5.7**   Notched box plots of formant frequencies: (a) F1, (b) F2, and (c) F3, computed using discrete all-pole modeling of apneic and benign snores for males (M), females (F), and both males and females combined (C). Threshold values are marked on the right-hand side of the plots.

OSA, and it can be a potential diagnostic marker for OSA.

Previous studies in speech and voice analysis have demonstrated that F1 is associated with the degree of constriction in the pharynx; F2 is related to the degree of advancement of the tongue relative to its neutral position; and F3 is correlated with the degree of lip-rounding [152,160,197-199]. For instance, a study in acoustic characteristics of voice after the UA surgery for snoring has shown a significant reduction in F1 of the vowel /i/ and in F2 of the vowel /o/ and /u/ attributable to the enlargement of the oropharyngeal cavity volume resulted from the anterior-posterior advancement of the tongue and the widening of the palatal arch caused by the uvulopalatopharyngoplasty and/or the tonsillectomy [197].

In this research work, the F1 of AS is significantly higher than of BS for same- and both-gender patient groups (p-value < 0.0001). Relating this to the pathophysiology of OSA, it is recognized that a higher AHI is associated with more collapsible pharynx [6,7], which is consistent with a higher degree of constriction of the UA, and thus, the increase in F1 of the snores. On the contrary, F2 and F3, which are related to the degree of advancement of the tongue and rounding of the lips, respectively [152,160,197-199], are noted to be not critical in the classification of apneic and benign patients.

To further appraise the diagnostic power of F1, we comparatively evaluated its performance with those obtained from the spectral PF – an extensively adopted indicator for identifying apneic and benign subjects [93-96], as discussed in Section 2.3, and to locate snoring sites [200-203]. For a fair test, PF was extracted by means of peak picking the PSD spectrum produced by the Welch's averaged modified periodogram method [112,113] under the same experimental conditions (900 AS and 300 BS for training, 300 AS and 100 BS for testing, and a 256-sample Hanning

**Table 5.3**   Diagnostic performance of spectral peak frequency (PF) computed using the Welch's method (frame size = 256 samples or 23 ms) of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
|------|-----|-----------|-----------|-----|-----|----------|----------|---------|
| M:A | 52.8 ± 24.7 | 652 ± 417 (97–2341) | | | | | | |
| M:B | 5.3 ± 3.5 | 287 ± 175 (96–1041) | 336 | 0.8122 | 0.0125 | 71.7 | 70.0 | < 0.0001 |
| F:A | 23.6 ± 15.1 | 577 ± 556 (99–2600) | | | | | | |
| F:B | 3.4 ± 3.2 | 211 ± 218 (88–1753) | 220 | 0.9091 | 0.0082 | 91.7 | 97.5 | < 0.0001 |
| C:A | 46.9 ± 25.7 | 637 ± 449 (97–2600) | | | | | | |
| C:B | 4.6 ± 3.4 | 257 ± 196 (88–1753) | 275 | 0.8411 | 0.0114 | 79.7 | 72.0 | < 0.0001 |

Values are presented as mean ± standard deviation and range within brackets. AHI refers to apnea-hypopnea index in events/h; Thre, threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

window with 75% overlap). Results in Table 5.3 confirm that F1 can deliver better diagnostic accuracy than PF. For both genders combined, the threshold of PF = 275 Hz, with a resultant AUC = 0.8411, SE = 0.0114, sensitivity = 79.7%, specificity = 72.0%, and p-value < 0.0001. In all three patient groups, the thresholds of PF attain a sensitivity of 71.7–91.7% and a specificity of 70.0–97.5%.

## 5.7    Diagnostic Marker Regression Models

To illustrate the possible relationship between AHI and F1, we attempted various types of regression models [204-207] for fitting the scatter plot of AHI versus F1. These models include linear, quadratic, logarithmic, exponential, and power functions, as rendered in Figures 5.8 and 5.9. Parameter estimates for the regression equations, as well as goodness-of-fit statistics, such as coefficient of determination ($R^2$) and residual
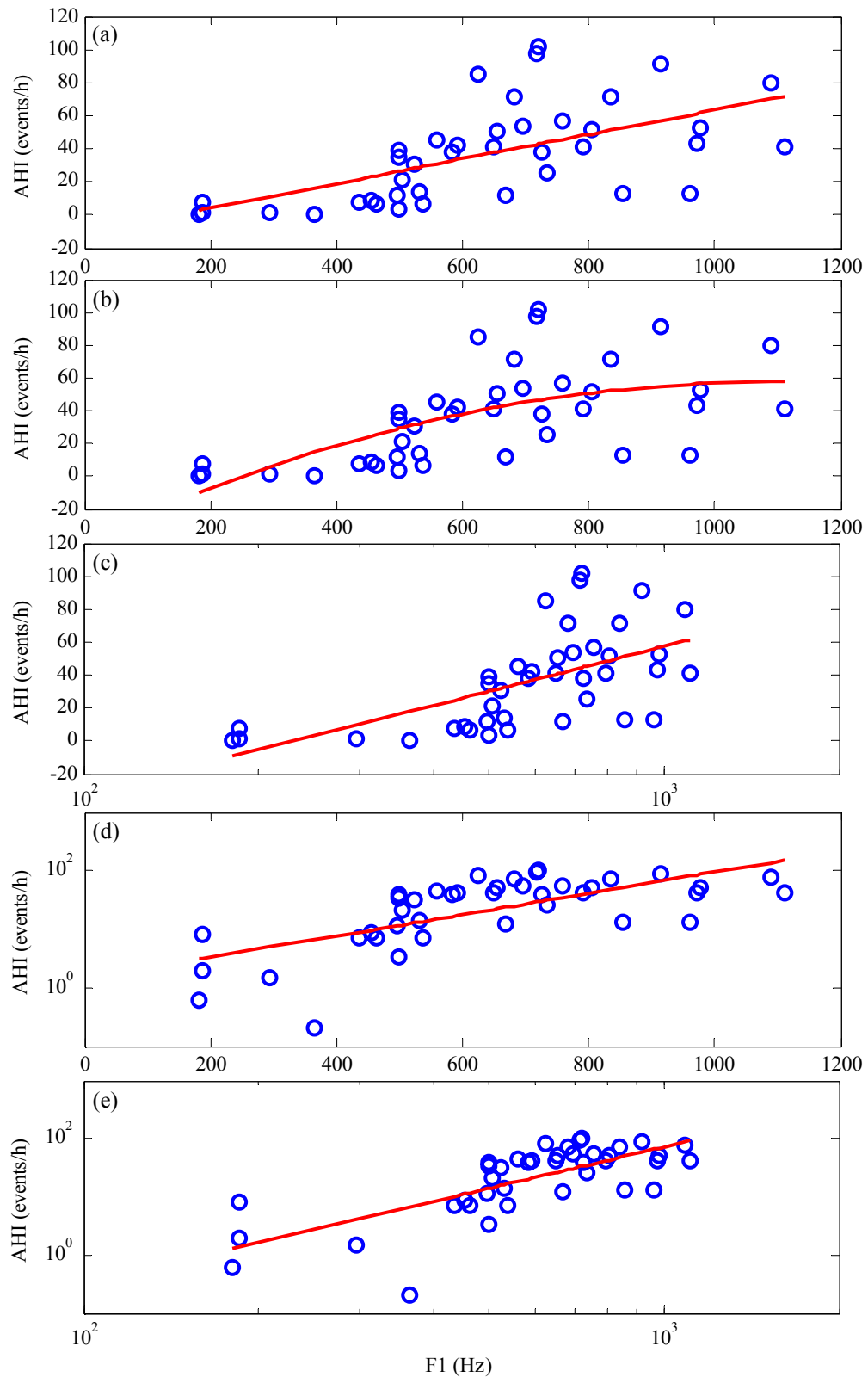
**Figure 5.8**   Regression models of (a) linear, (b) quadratic, (c) logarithmic, (d) exponential, and (e) power functions for apnea-hypopnea index (AHI) in events/h and first formant frequency (F1) computed using linear prediction in hertz (Hz).
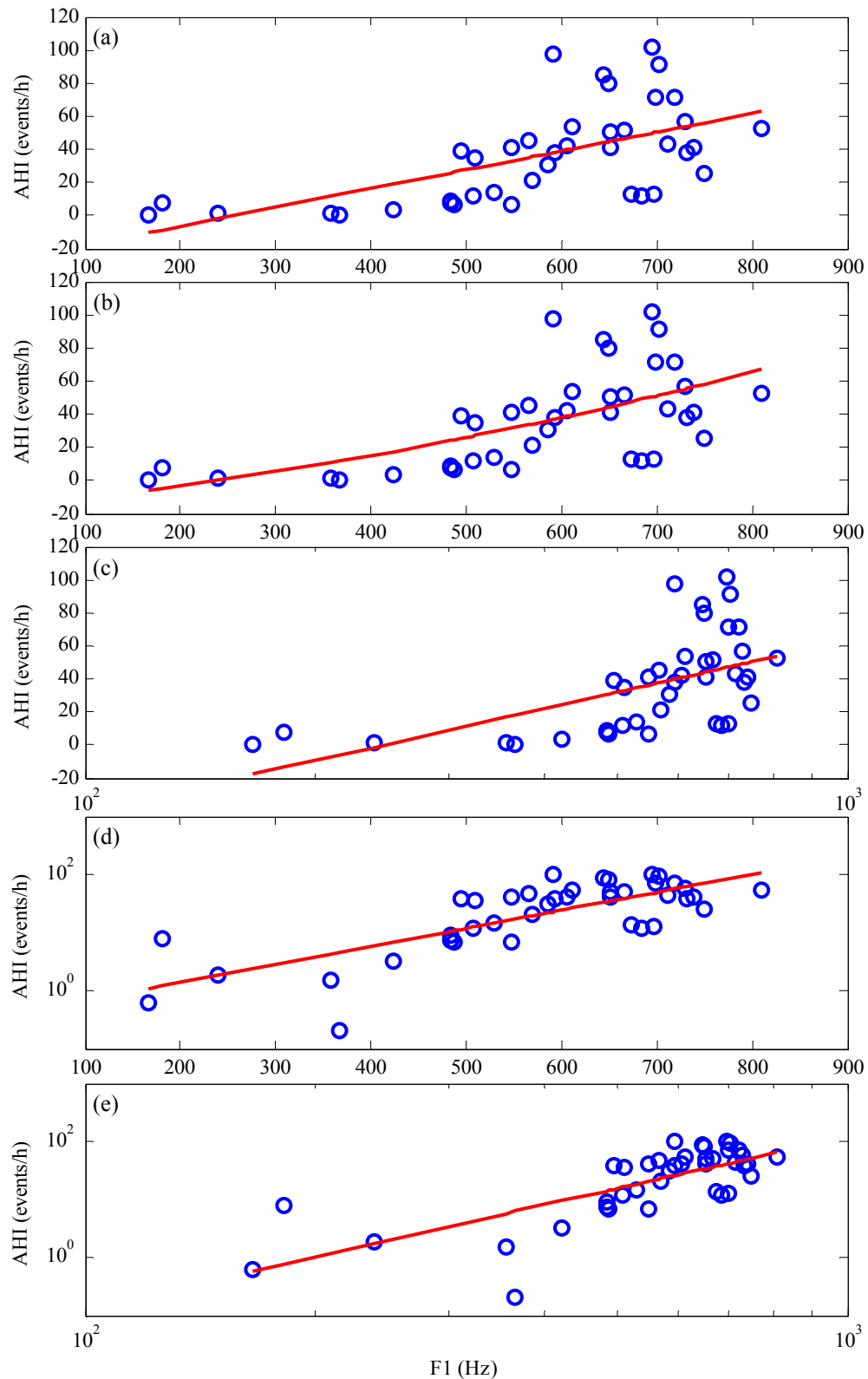
**Figure 5.9**   Regression models of (a) linear, (b) quadratic, (c) logarithmic, (d) exponential, and (e) power functions for apnea-hypopnea index (AHI) in events/h and first formant frequency (F1) computed using discrete all-pole modeling in hertz (Hz).

**Table 5.4**    Regression models for apnea-hypopnea index (AHI) in events/h and first formant frequency (F1) computed using linear prediction in hertz (Hz).

| Regression | Equation | $R^2$ | RSD | AHI[†] |
|---|---|---|---|---|
| Linear | AHI = 0.0738 (F1) – 10.4067 | 0.3449 | 23.7718 | 24.3 |
| Quadratic | AHI = – 0.0001 (F1)$^2$ + 0.1778 (F1) – 39.2095 | 0.3824 | 23.0811 | 26.3 |
| Logarithmic | AHI = – 212.4891 + 89.9464 Log (F1) | 0.3553 | 23.5832 | 27.9 |
| Exponential | Log (AHI) = 0.1545  + 0.0018 (F1) | 0.4602 | 0.4614 | 10.3 |
| Power | Log (AHI) = – 5.2100 + 2.3565 Log (F1) | 0.5334 | 0.4290 | 12.2 |

[†]Predictive AHI corresponding to derived thresholds for both males and females combined (F1 = 470 Hz). Ideally, AHI = 10.0 events/h. $R^2$ refers to coefficient of determination; RSD, residual standard deviation in events/h; Log, logarithm function with base 10.

**Table 5.5**    Regression models for apnea-hypopnea index (AHI) in events/h and first formant frequency (F1) computed using discrete all-pole modeling in hertz (Hz).

| Regression | Equation | $R^2$ | RSD | AHI[†] |
|---|---|---|---|---|
| Linear | AHI = 0.1138 (F1) – 29.2159 | 0.3580 | 23.5344 | 26.9 |
| Quadratic | AHI = 0.0001 (F1)$^2$ + 0.0571 (F1) – 16.8990 | 0.3613 | 23.4732 | 25.2 |
| Logarithmic | AHI = – 249.6060 + 104.3999 Log (F1) | 0.3119 | 24.3632 | 31.5 |
| Exponential | Log (AHI) = – 0.4728 + 0.0031 (F1) | 0.5784 | 0.4077 | 11.3 |
| Power | Log (AHI) = – 6.7838 + 2.9547 Log (F1) | 0.5465 | 0.4229 | 14.9 |

[†]Predictive AHI corresponding to derived thresholds for both males and females combined (F1 = 493 Hz). Ideally, AHI = 10.0 events/h. $R^2$ refers to coefficient of determination; RSD, residual standard deviation in events/h; Log, logarithm function with base 10.

standard deviation (RSD), are tabulated in Table 5.4 and Table 5.5 for F1 computed using LP and DAP modeling, respectively. The present equations are derived based on the mean F1 of the training data from each patient and their corresponding AHI. Based on each type of regression equation, one can surmise a predictive AHI corresponding to the F1 threshold of 470 Hz for LP and 493 Hz for DAP modeling.

Results suggest that the possible relationship between AHI and F1 can likely take the functional form of exponential (LP: $R^2$ = 0.4602, RSD = 0.4614 events/h, predicted AHI = 10.3 events/h; DAP: $R^2$ = 0.5784, RSD = 0.4077 events/h, predicted AHI = 11.3 events/h) or power (LP: $R^2$ = 0.5334, RSD = 0.4290 events/h, predicted

AHI = 12.2 events/h; DAP: $R^2$ = 0.5465, RSD = 0.4229 events/h, predicted AHI = 14.9 events/h). The corresponding predicted AHI values for these regression models are consistent with the AHI cutoff value of 10.0 events/h for AS and BS.

## 5.8    Summary

This chapter introduces a parametric approach via LP and DAP modeling to investigate the feasibility of using formant frequencies of snore signals for diagnosing OSA. Quantitative differences in formant frequencies between AS and BS are found in same- and both-gender patient groups, lending support to our hypothesis that snore signals of apneic and benign patients may have distinguishable formant features due to differences in their UA shapes and physical dimensions. AS exhibit higher formant frequencies than BS, especially F1, which can be related to the pathophysiology of OSA. Regardless of the techniques (LP or DAP modeling) used for computing F1, the diagnostic performance of F1 for all three patient groups is superior to that of PF, with a sensitivity of 82.5–100% versus 71.7–91.7% and a specificity of 82.0–95.0% versus 70.0–97.5%. Furthermore, the relationship between AHI and F1 can best describe by the exponential or the power regression model, with a predictive AHI of 10.3–14.9 events/h.

While the parametric approach of analyzing snore signals has shed light on the detection of OSA, it is tailored to the framework of classical linear models and ergodic Gaussian random processes. Consequently, it may not adequately characterize snore signals that are generally nonstationary, nonlinear, and non-Gaussian in nature [99,149,208-212]. The next chapter of this thesis attempts to study and assess nonlinearities in snores using a wavelet-driven nonlinear approach, and subsequently, classify the snores generated by apneic and benign patients.

Chapter 6

# Nonlinear Analysis and Classification of Snores

Linear approximations are frequently adopted to model nonlinear phenomena due to their simplicity. Mathematical and structural models of the UA as a collapsible tube for studying snore production have been introduced according to linear acoustic theory [28,150,213-216]. Although these models shed light on the pathology of snoring, there are strong evidences for the occurrence of nonlinear fluid dynamics (airflow) phenomena during snoring that cannot be easily represented by linear models [149,208-212]. The UA is a highly complex, three-dimensional system involving irregular geometric structures and tissues with different biomechanical properties and functions. Soft tissues are generally governed by nonlinear viscoelastic behavior. Owing to their nonlinear properties, interactions between fundamental modes of oscillation in snore production can lead to certain phase relations among frequency components [146,162], just like those in speech production [61-67,217,218]. Hence, we hypothesize that phase-coupled interaction modes may usefully classify AS from BS since patients with OSA have anatomical and functional abnormalities associated with their UA [29-31,56,57].

This chapter examines the feasibility of employing a nonlinear dynamics perspective in the analysis of snore signals by means of wavelet bicoherence (WBC). Quadratic phase coupling (QPC), a signature of nonlinear interactions, between frequency modes in snore signals is highlighted. Two novel diagnostic markers, namely peak frequency at $f_1$ (PF1) and peak sum frequency (PSF) in their respective

_____

The contents of this chapter are partly based on the author's publications [A3,A8].

WBC spectrum, are proposed to discriminate AS from BS in same- and both-gender patient groups. In the same fashion as Chapter 5, we comprehensively presented and discussed the experimental considerations prior to signal analysis, diagnostic performance of the proposed markers, and regression models between AHI and the proposed markers.

## 6.1    Wavelet Higher-Order Statistics

Wavelet polyspectral techniques have received wide acceptance in the analysis of nonstationary and nonlinear signals because of their ability to simultaneously decompose a signal into both time and frequency, unveiling possible evolutionary phenomena hidden in the signal. The $p$th-order wavelet polyspectral [219,220] relative to an admissible wavelet

$$\psi(t) = g(t)\exp(j\eta t) \tag{6.1}$$

where $g(t)$ is a real-valued symmetric function, $\eta$ is the central frequency of the wavelet, and $j = \sqrt{-1}$ can be defined as

$$M_p\left(a_1,\ldots,a_{p-1};t_0\right) = \frac{1}{T}E\left\{\int_{t_0-T/2}^{t_0+T/2} W_\psi^*(t,a_0)W_\psi(t,a_1)\ldots W_\psi(t,a_{p-1})dt\right\}. \tag{6.2}$$

$$W_\psi(t,a_0) = \frac{1}{\sqrt{a_0}}\int_{-\infty}^{\infty} x(t')\psi^*\left(\frac{t'-t}{a_0}\right)dt' \tag{6.3}$$

is the continuous wavelet transform (CWT) [136,221] of a time series signal $x(t)$ whose process over an interval $[t_0\text{-}T/2,\ t_0\text{+}T/2]$ at time instant $t_0$ for localized time window $T$ is assumed to be quasi-stationary, with an inverse sum rule

$$a_0^{-1} = a_1^{-1} + a_2^{-1} + \ldots + a_{p-1}^{-1} \tag{6.4}$$

that establishes a correlation between scale $a_0$ and instantaneous frequency

$$f_0 = \frac{\eta}{2\pi a_0}.$$ 

(6.5)

Consider a continuous-time denoised snore signal $x(t) = x(n\Delta t)$ with $n$ as an integer referring to the sample number and $\Delta t$ as the signal sampling period, a time-discretized $p$th-order wavelet polyspectral [219,220] at $t_0 = L\Delta t$ for an interval [($L-K)\Delta t$, $(L+K)\Delta t$] can be formulated as

$$\hat{M}_p(f_1, \ldots, f_{p-1}; L\Delta t) = \frac{1}{2K+1} \sum_{k=L-K}^{L+K} \hat{W}_\psi^*(k, f_0)\hat{W}_\psi(k, f_1)\ldots\hat{W}_\psi(k, f_{p-1})$$ 

(6.6)

where

$$\hat{W}_\psi(k, f_0) = \sqrt{\frac{2\pi f_0}{\eta}} \Delta t \sum_{n=0}^{N-1} x(n)\psi^*\left[\frac{2\pi f_0}{\eta}(n-k)\Delta t\right]$$ 

(6.7)

is the CWT estimator, $k \in \mathbb{N}$, $K = T/(2\Delta t)$, and

$$f_0 = f_1 + f_2 + \ldots + f_{p-1}.$$ 

(6.8)

## 6.2     Wavelet Bicoherence

### 6.2.1     Quadratic Phase Coupling

The normalized 3rd-order wavelet polyspectrum, also known as WBC, offers a quantitative measure of the presence and strength of QPC induced by nonlinearities in the signal production mechanism [219,220,222]. It encompasses both wavelet and higher-order statistics for monitoring phase coupling and nonlinear interactions. The robustness of WBC analysis has been demonstrated in previous studies of wind-wave and wave-wave interactions [223-225], dynamics of inhomogeneous electron-plasma systems [226,227], as well as characteristics of asthmatic wheezes [228].

Based on the earlier discussion in Equation (6.6), the squared WBC can be expressed as

$$\hat{b}^2_{\text{WBC}}(f_1, f_2; L\Delta t) = \frac{\left|\hat{M}_3(f_1, f_2; L\Delta t)\right|^2}{\frac{1}{2K+1}\sum_{k=L-K}^{L+K}\left|\hat{W}_\psi(k, f_1)\hat{W}_\psi(k, f_2)\right|^2 \hat{M}_2(f_1 + f_2; L\Delta t)} \tag{6.9}$$

where

$$\hat{M}_2(f_1 + f_2; L\Delta t) = \hat{M}_2(f_0; L\Delta t) = \frac{1}{2K+1}\sum_{k=L-K}^{L+K}\left|\hat{W}_\psi(k, f_0)\right|^2 \tag{6.10}$$

and

$$\hat{M}_3(f_1, f_2; L\Delta t) = \frac{1}{2K+1}\sum_{k=L-K}^{L+K}\hat{W}^*_\psi(k, f_0)\hat{W}_\psi(k, f_1)\hat{W}_\psi(k, f_2) \tag{6.11}$$

symbolize the wavelet power spectrum and the wavelet bispectrum, correspondingly.

QPC is often characterized to occur when two frequency modes at $f_1$ and $f_2$ with their corresponding phases $\theta_1$ and $\theta_2$ are concurrently present in a signal with a frequency at $f = f_1 \pm f_2$ with $\theta = \theta_1 \pm \theta_2$. The value of WBC is limited between 0 and 1, where 0 indicates absence of QPC, and 1 indicates complete occurrence of QPC between $f_1$ and $f_2$. One of the key advantages of WBC over Fourier-based bicoherence is that it can better identify short-lived quadratic interactions for both with and without phase coupling [222,229].

To exemplify the concept on QPC, consider the UA as a quadratic acoustic system of the form

$$x(t) = u^2(t) + u(t) \tag{6.12}$$

where $x(t)$ is the output of the system, and

$$u(t) = A_1 \cos(w_1 t + \theta_1) + A_2 \cos(w_2 t + \theta_2) \tag{6.13}$$

is the input to the system with $A_i$, $w_i = 2\pi f_i$, and $\theta_i$ as the amplitude, angular frequency, and phase of the $i$th excitation source waves, respectively. Then,

$$
\begin{aligned}
x(t) = {} & 0.5\left(A_1^2 + A_2^2\right) + A_1 \cos\left(w_1 t + \theta_1\right) + A_2 \cos\left(w_2 t + \theta_2\right) \\
& + A_1 A_2 \cos\left(w_3 t + \theta_3\right) + A_1 A_2 \cos\left(w_4 t + \theta_4\right) \\
& + 0.5 A_1^2 \cos\left(2 w_1 t + 2\theta_1\right) + 0.5 A_2^2 \cos\left(2 w_2 t + 2\theta_2\right)
\end{aligned}
\tag{6.14}
$$

where $w_3 = w_1 + w_2$, $w_4 = w_1 - w_2$, $\theta_3 = \theta_1 + \theta_2$, and $\theta_4 = \theta_1 - \theta_2$.

Apparently, the system output contains the input frequency components, harmonics at $(2w_1, 2\theta_1)$ and $(2w_2, 2\theta_2)$, as well as intermodulation components at $(w_3, \theta_3)$ and $(w_4, \theta_4)$, which arise from nonlinear interactions between the primary frequency modes, $w_1$ and $w_2$, or equivalently, $f_1$ and $f_2$. Thus, applying WBC analysis to the output will yield a peak, whose value is close to unity, at the bifrequency $(f_1, f_2)$. On the other hand, in the absence of nonlinearities within the acoustic system, there will be no phase-coupled interactions between the fundamental modes. Consequently, no QPC peak will exist at the bifrequency; the corresponding WBC value will be approximately zero.

## 6.2.2    Feature Extraction

A squared WBC spectrum measures the fraction of power at a given frequency attributed to nonlinear mode interactions [219,220,222]. To facilitate easy interpretation, the squared WBC is normally plotted on a bifrequency $(f_1, f_2)$ plane shown in Figure 6.1 using its values in the principal domain, which comprises of an inner triangle ($0 \leq f_1 \leq$ Nyquist frequency $f_N$, $0 \leq f_2 \leq f_N$, $0 \leq f_1 + f_2 \leq f_N$) and an outer triangle ($0 \leq f_2 \leq f_1, f_N \leq f_1 + f_2 \leq 2 f_N - f_1$) [222,230].

Two metrics are proposed to distinguish between AS and BS. Both metrics rely on the summation of squared WBC in the principle domain: one measures the relative coupling strength contributed by the frequency mode at $f_1$,

$$
b^{\mathrm{PF1}}\left(f_1\right) = \sum_{f_1} \hat{b}_{\mathrm{WBC}}^2\left(f_1, f_2\right),
\tag{6.15}
$$
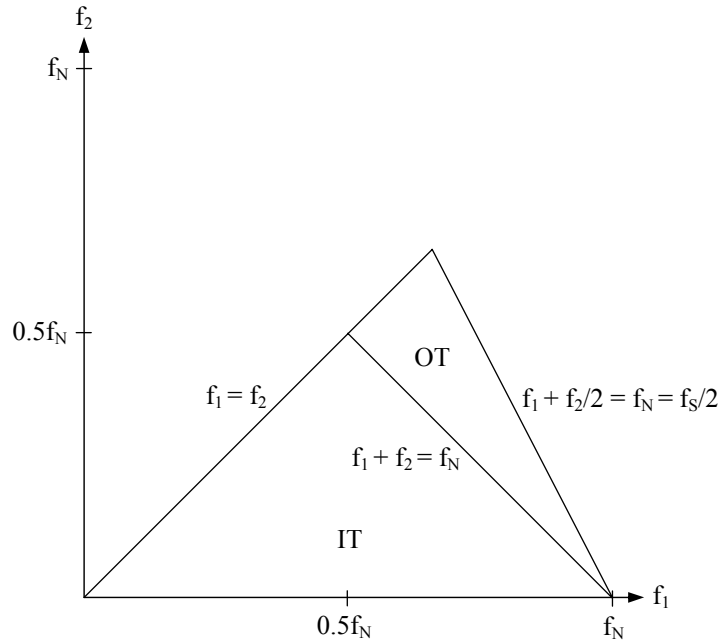
**Figure 6.1**    A bifrequency ($f_1$, $f_2$) plane with principal domain comprising an inner triangle (IT) and an outer triangle (OT). $f_N$ and $f_S$ denote Nyquist frequency and sampling frequency, respectively.

whereas the other measures the total coupling strength to a sum frequency $f = f_1 + f_2$ owing to interactive coupling at $f_1$ and $f_2$,

$$b^{\text{PSF}}(f) = \sum_{f=f_1+f_2} \hat{b}^2_{\text{WBC}}(f_1, f_2). \tag{6.16}$$

Frequency component that possesses the highest coupling strength in Equation (6.15) and in Equation (6.16) was extracted by peak peaking and denoted as PF1 and PSF, respectively.

Because wavelet coefficients of nonorthogonal wavelets (e.g., Morlet-Grossman) are not statistically independent, they may give rise to statistical noise level [222,228] in the estimation of WBC, with an upper limit given by

$$\varepsilon\left[\hat{b}^2_{\text{WBC}}(f_1, f_2)\right] = \frac{f_S}{2(2K+1)\min\left(|f_1|, |f_2|, |f_1+f_2|\right)}. \tag{6.17}$$

By inspecting the Equation (6.17), one sees that the statistical noise level is frequency

dependent and dominates the bicoherence at low frequencies but declines rapidly with increasing frequency. For reliability of this research work, we determined PF1 and PSF only when their corresponding metrics were above the noise level.

## 6.3    Experimental Conditions

### 6.3.1    Patient Dataset

The patient dataset is exactly the same as those described in Section 5.5.1. In a nutshell, snoring sounds were captured from 40 patients (30 apneic and 10 benign) tabulated in Table 1.1 via the snore signal acquisition system (see Chapter 3), and subsequently preprocessed using the preprocessing system within a TIDWT domain (see Chapter 4). From each patient, 40 inspiratory snores (30 as training and 10 as test data) were evaluated. Diagnostic accuracy of the proposed markers was quantified by means of ROC curves [155-157] and notched box plots [139,140,181,182], with a statistical significance of p-value < 0.05.

### 6.3.2    Optimal Parameter Selection

In the analysis of WBC, we selected the Morlet-Grossman wavelet

$$\psi(t) = \exp\left(-\frac{t^2}{2\sigma^2}\right)\exp(j\eta t) \tag{6.18}$$

as the mother wavelet due to its excellent localization properties in both time and frequency [136,219-221]. Even though this wavelet is not strictly admissible since the average of its function over the entire space is not equal to zero, i.e.,

$$\int_{-\infty}^{\infty}\psi(t)dt \neq 0 \,, \tag{6.19}$$

appropriate considerations of the wavelet standard deviation $\sigma^2$ and central frequency $\eta$
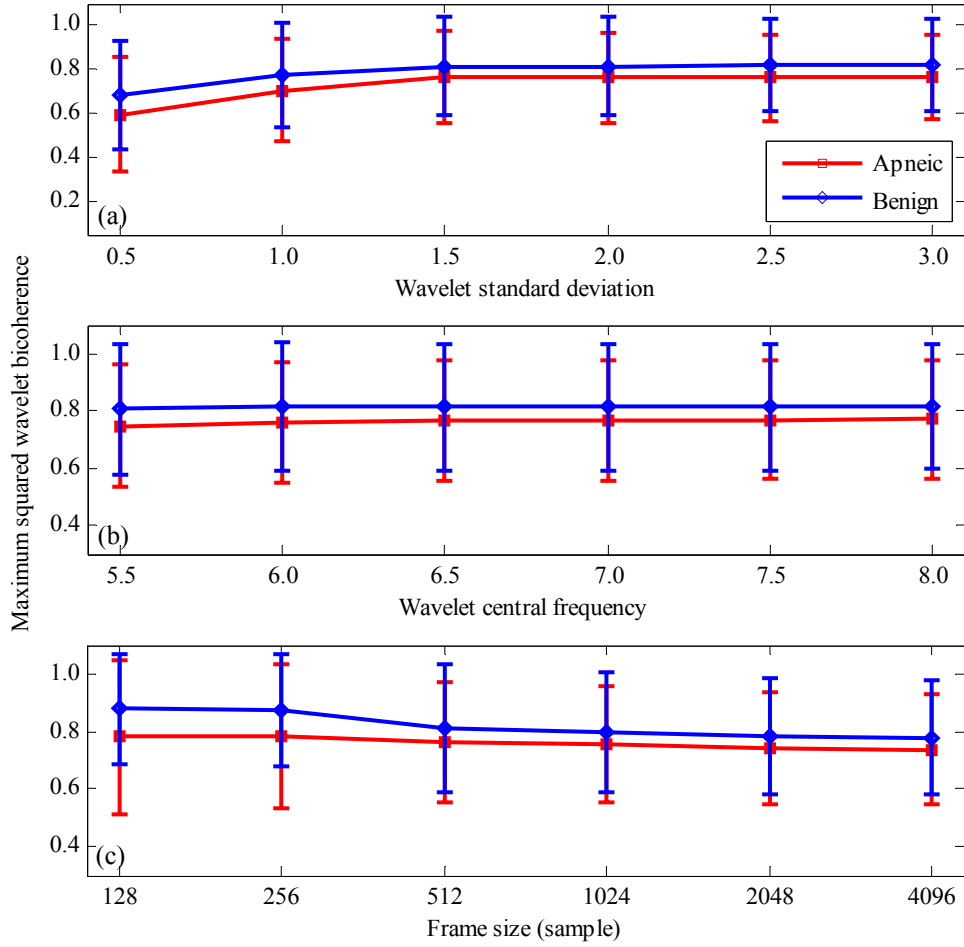
**Figure 6.2**     Maximum magnitude of squared wavelet bicoherence for different values of (a) wavelet standard deviation $\sigma^2$ at wavelet central frequency $\eta = 6$, (b) $\eta$ at $\sigma^2 = 1.5$, and (c) frame size at $\sigma^2 = 1.5$ and $\eta = 6$ for apneic and benign snores. Error bars indicate one standard deviation.

can yield $\psi(0) \approx 0$ [136,231]. For instance, if $\sigma^2 = 1$ and $\eta \geq \pi(2/\log_e 2)^{1/2} \approx 5.3364$, then $\psi(0) < 10^{-33}$, which can be deemed as numerically admissible. In this work, we employed $\sigma^2 = 1.5$ and $\eta = 6$. Figures 6.2a and 6.2b highlight the maximum magnitude of squared WBC in Equation (6.9), which was computed at a time instant $t_0$ located at the center of each snore segment, for each possible value of the standard deviation ($\sigma^2 = 0.5: 0.5: 3$) and central frequency ($\eta = 5.5: 0.5: 8$). It can be noted that the magnitude levels almost settle down at $\sigma^2 = 1.5$ and $\eta = 6$ for both AS and BS. Furthermore, these recommended wavelet parameters satisfy the condition of admissibility.

Besides tuning the wavelet parameters for better localization, the selection of frame size of snore signals is equally crucial for WBC analysis because this parameter recognizes the competence of the estimated WBC to account for alterations in the signal properties [228]. Similar to the earlier computational procedures, Figure 6.2c depicts the maximum magnitude of squared WBC for various frame sizes (128, 256, 512, 1024, and 2048 samples, or equivalently, about 12 ms, 23 ms, 46 ms, 93 ms, and 186 ms). As can be seen, the magnitude fluctuation of squared WBC for AS and BS are rarely significant as the frame size increases beyond 512 samples, implying a favorable frame size of 512 samples ($\approx$ 46 ms) for the analysis.

## 6.4    Diagnostic Results and Discussion

For an illustrative example of WBC analysis of snore signals, Figure 6.3 presents an approximated magnitude of the squared CWT of a typical apneic snore and benign snore, along with its temporal counterpart. It is apparent from the plots that spectral peaks of the apneic snore sweep across a larger frequency band (200–1200 Hz) in an unpredictable manner, while the peaks of the benign snore uniformly focus on a smaller band (50–250 Hz). The discrepancy in the time-frequency properties of snores accentuates the importance of wavelet polyspectral analysis in detecting short-lived nonlinear interactions that are veiled by the stationary assumption of the Fourier-based approach.

Figure 6.4 illustrates the squared WBC spectra, $b^{\mathrm{PF1}}(f_1)$, and $b^{\mathrm{PSF}}(f)$ of a frame size of the same snores. Apart from a wider frequency band, the apneic snore exhibits nonlinear couplings in higher frequency range (200–1000 Hz) than that of the benign one (50–500 Hz), as evidently displayed on the contour plots in Figures 6.4a and 6.4b. Explicitly, the QPC peaks of the apneic and the benign snore are located at $(f_1, f_2) \approx$
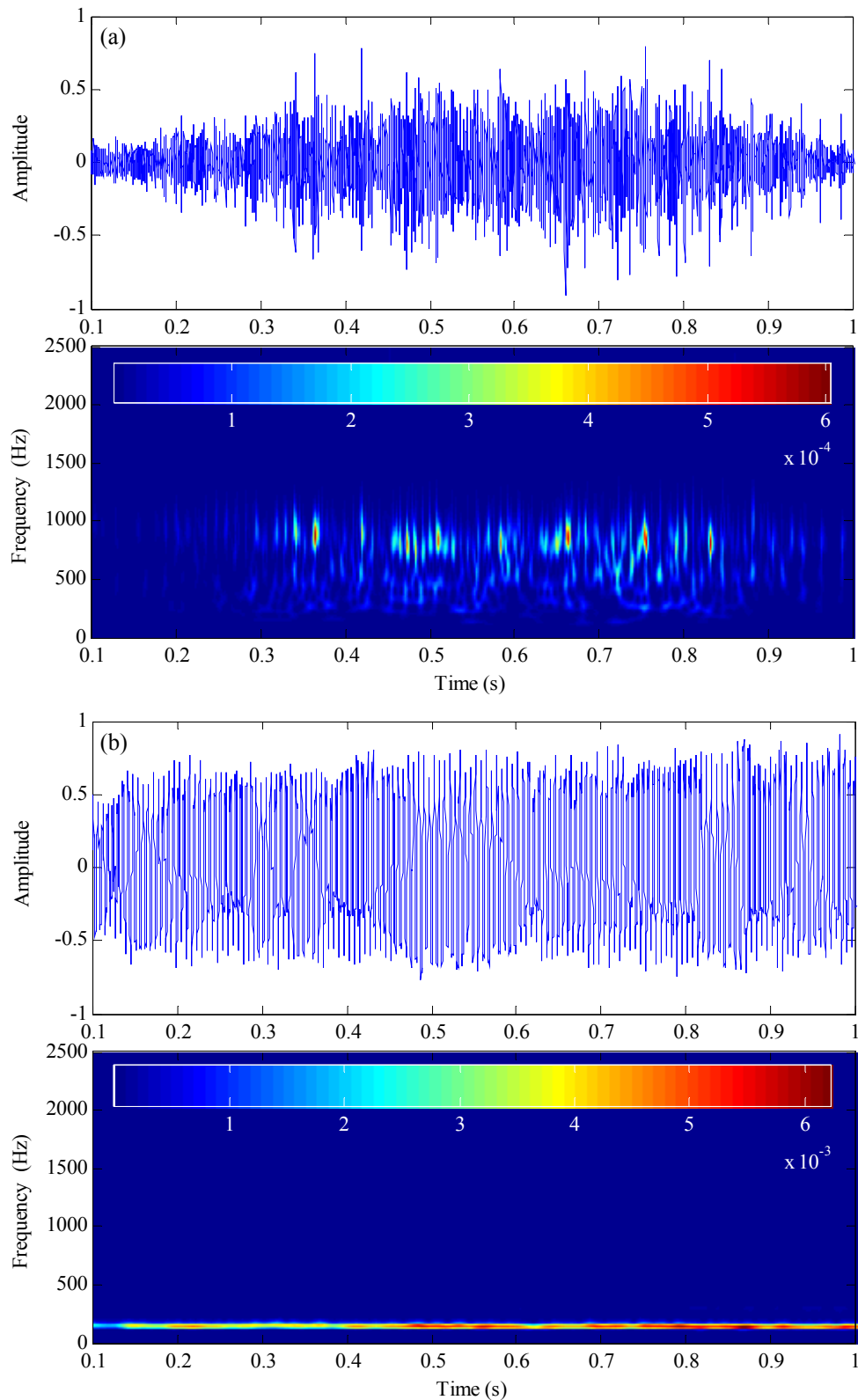
**Figure 6.3**    Temporal and wavelet time-frequency features of a typical (a) apneic snore and (b) benign snore.
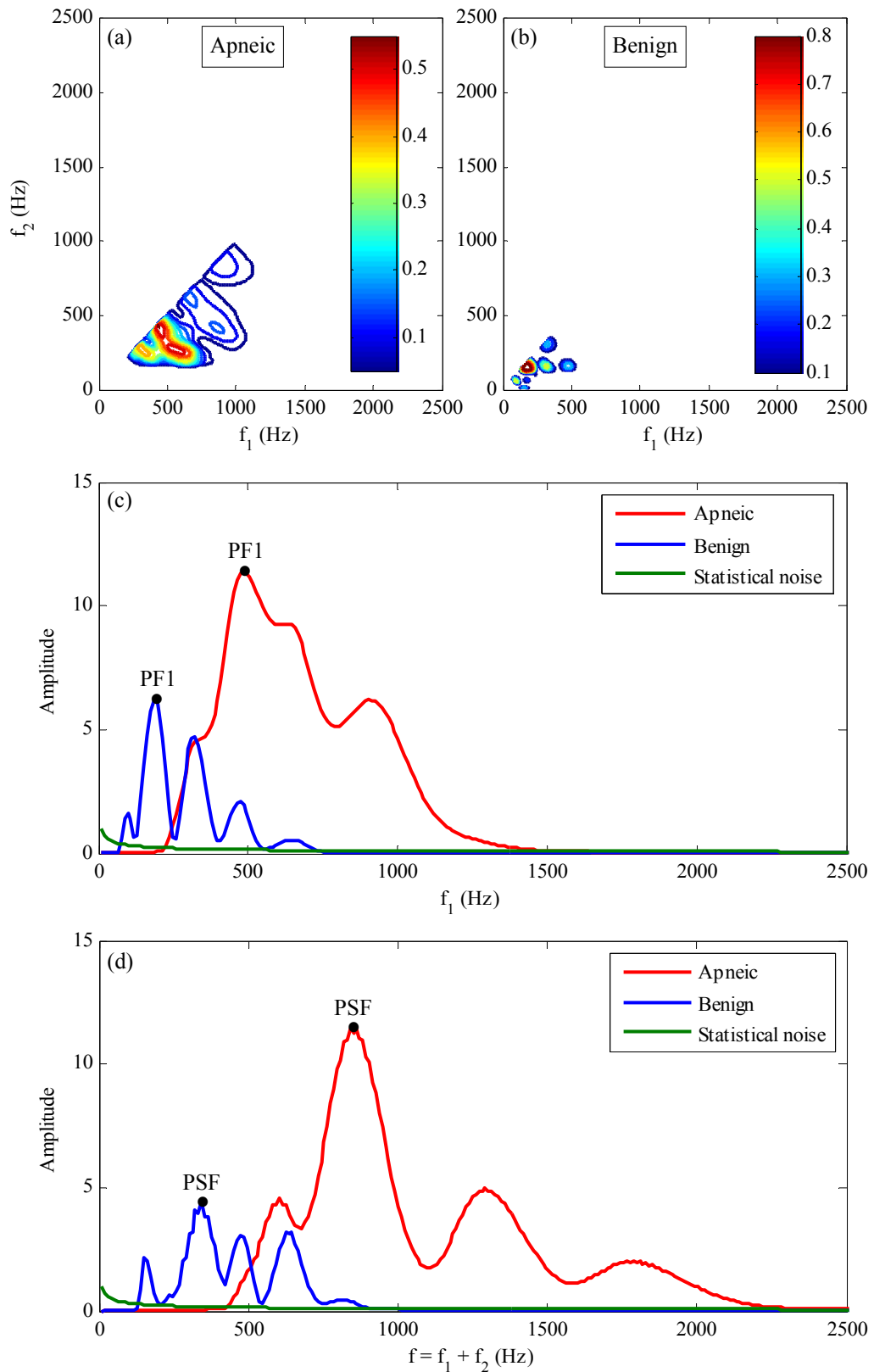
**Figure 6.4**   Squared wavelet bicoherence spectra of a typical (a) apneic snore and (b) benign snore, and their corresponding (c) $b^{PF1}(f_1)$ and (d) $b^{PSF}(f)$ plots. Dots in (c) and (d) denote the peaks of PF1 and PSF, respectively.

(441, 419) Hz with a sum frequency at $f_{AS} = f_1 + f_2 \approx 860$ Hz, and $(f_1, f_2) \approx (172, 162)$ Hz with a sum frequency at $f_{BS} = f_1 + f_2 \approx 334$ Hz, respectively. A close inspection of the contour plots justifies the existence of $f_{AS}$ and $f_{BS}$. Moreover, by examining the WBC analytical outcome of each frame of a snore in the existing dataset, we noticed that QPC can occur over different bands of frequencies, and this band can differ from one frame to another, in particular for AS, as evidenced in the CWT plots. As discussed in Section 1.1.1, patients harboring of OSA experience recurrent episodes of UA narrowing and collapse during sleep. The instable UA, together with structural anomalies and anatomical lesions, often generate loud chaotic snoring sounds from increased turbulence of airflow, which may lead to complex dynamic extension of frequency patterns in nonlinear interactions.

To facilitate easy comparisons between the representative apneic and benign snore in the degree of nonlinearity, we qualitatively summarized the above WBC information by the metrics, $b^{PF1}(f_1)$ and $b^{PSF}(f)$, as rendered in Figures 6.4c and 6.4d, respectively. Both plots consistently suggest that mode-mode interactions in the apneic snore are spanned in the region of higher frequencies between different components, whereas interactions in the benign one are in lower region with high self-coupling ($f_1 = f_2$) tendency, thereby manifesting themselves as distinctive peaks that resemble harmonic series. Accordingly, the proposed markers of the snores are PF1(apneic; benign) $\approx$ (485; 183) Hz, and PSF(apneic; benign) $\approx$ (861; 345) Hz, which are closely related to the earlier findings. These markers are good estimations of the frequency modes that extensively contribute in the nonlinear phase coupling, and they are reliable as their corresponding metrics are much above the statistical noise level.

Further analysis of the dataset shows that AS are roughly 20% less self-coupled than BS. This can be elucidated by the fact that AS are commonly composed of non-

rhythmic, high-frequency patterns with spectral PF above 400 Hz attributable to the vibrations of tissues at tongue base level, which is a prevailing occlusion site for OSA, and/or at other non-palatal snoring sites (e.g., epiglottis and supraglottic larynx), accompanied by turbulent noise. Conversely, BS are likely originated from soft palate and/or uvula, resulting in simple, low-frequency harmonics with PF below 400 Hz [93-95]. The PF of snores produced from soft palate, tonsils, epiglottis, and base of tongue can lie in the range of 105–189 Hz, 85–201 Hz, 331–510 Hz, and 1215–1277 Hz, respectively [201]; the fundamental frequency of palatal snores and tongue base snores can be of $103 \pm 35$ Hz and $332 \pm 145$ Hz, respectively  [202].

Clinical usefulness of PF1 and PSF were also assessed. Diagnostic outcomes contained in Table 6.1 and Table 6.2 reveal that PF1 and PSF can potentially serve as diagnostic markers to distinguish apneic patients from benign ones, with significant quantitative differences between AS and BS (p-value < 0.0001) for all three patient groups (males, females, and combined). For both genders combined, the thresholds of PF1 = 285 Hz and PSF = 492 Hz yield resultant AUC (PF1; PSF) = (0.9694; 0.9652), SE (PF1; PSF) = (0.0044; 0.0047), and p-value (PF1; PSF) = (< 0.0001; < 0.0001). These threshold values fall between those derived from the same gender patients, being higher for males (PF1 = 325 Hz; PSF = 551 Hz) than females (PF1 = 210 Hz; PSF = 368 Hz).

A probable contributing factor for females having relative small threshold values in their respective groupings, apart from the modest sample size of 6 apneic and 4 benign female subjects, is the elevated degree of UA stability in females than males due to either better UA anatomy or increased UA dilator muscle activity [192-195]. This is also the rationale why the female UA is more susceptible to partial than complete obstruction [196]. Nonetheless, in all three groupings, the thresholds of PF1

**Table 6.1**    Diagnostic performance of peak frequency at $f_1$ (PF1) computed using wavelet bicoherence analysis of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
|------|-----|-----------|-----------|-----|-----|---------|---------|---------|
| M:A | 52.8 ± 24.7 | 908 ± 508 (170–2347) | | | | | | |
| M:B | 5.3 ± 3.5 | 256 ± 112 (139–705) | 325 | 0.9501 | 0.0058 | 85.4 | 85.0 | < 0.0001 |
| F:A | 23.6 ± 15.1 | 765 ± 487 (235–2302) | | | | | | |
| F:B | 3.4 ± 3.2 | 164 ± 24 (110–210) | 210 | 1.0000 | 0.0000 | 98.3 | 100 | < 0.0001 |
| C:A | 46.9 ± 25.7 | 880 ± 507 (170–2347) | | | | | | |
| C:B | 4.6 ± 3.4 | 219 ± 99 (110–705) | 285 | 0.9694 | 0.0044 | 90.7 | 85.0 | < 0.0001 |

Values are presented as mean ± standard deviation and range within brackets. AHI refers to apnea-hypopnea index in events/h; Thre, threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

**Table 6.2**    Diagnostic performance of peak sum frequency (PSF) computed using wavelet bicoherence analysis of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
|------|-----|-----------|-----------|-----|-----|---------|---------|---------|
| M:A | 52.8 ± 24.7 | 1334 ± 715 (249–3439) | | | | | | |
| M:B | 5.3 ± 3.5 | 432 ± 193 (233–1265) | 551 | 0.9431 | 0.0062 | 82.5 | 83.3 | < 0.0001 |
| F:A | 23.6 ± 15.1 | 1253 ± 755 (363–3358) | | | | | | |
| F:B | 3.4 ± 3.2 | 256 ± 53 (172–425) | 368 | 0.9998 | 0.0004 | 100 | 100 | < 0.0001 |
| C:A | 46.9 ± 25.7 | 1318 ± 723 (249–3439) | | | | | | |
| C:B | 4.6 ± 3.4 | 361 ± 176 (172–1265) | 492 | 0.9652 | 0.0047 | 90.3 | 86.0 | < 0.0001 |

Values are presented as mean ± standard deviation and range within brackets. AHI refers to apnea-hypopnea index in events/h; Thre, threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.
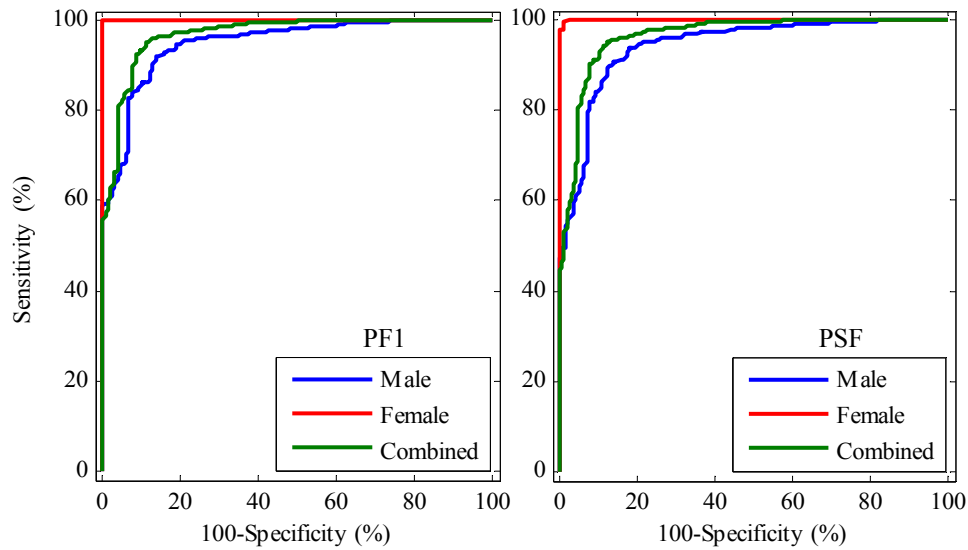
**Figure 6.5**   Receiver operating characteristic curves of peak frequency at $f_1$ (PF1) and peak sum frequency (PSF) of apneic and benign snores for males, females, and both males and females combined.
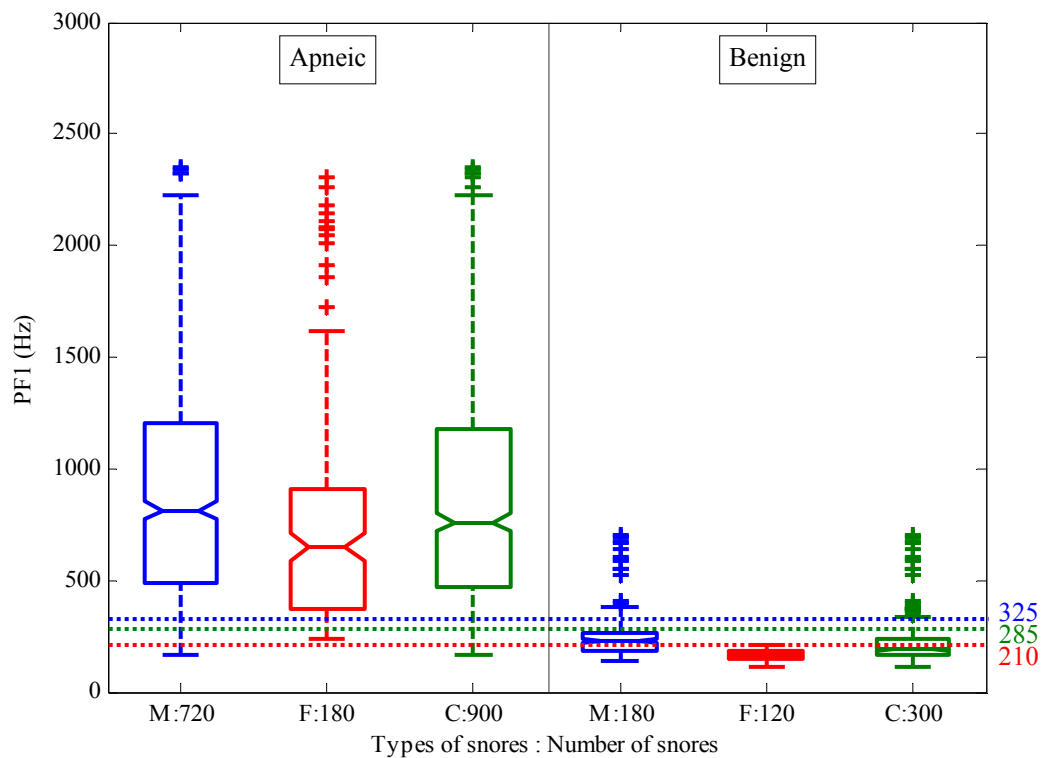


**Figure 6.6**   Notched box plots of peak frequency at $f_1$ (PF1) of apneic and benign snores for males (M), females (F), and both males and females combined (C). Threshold values are marked on the right-hand side of the plots.
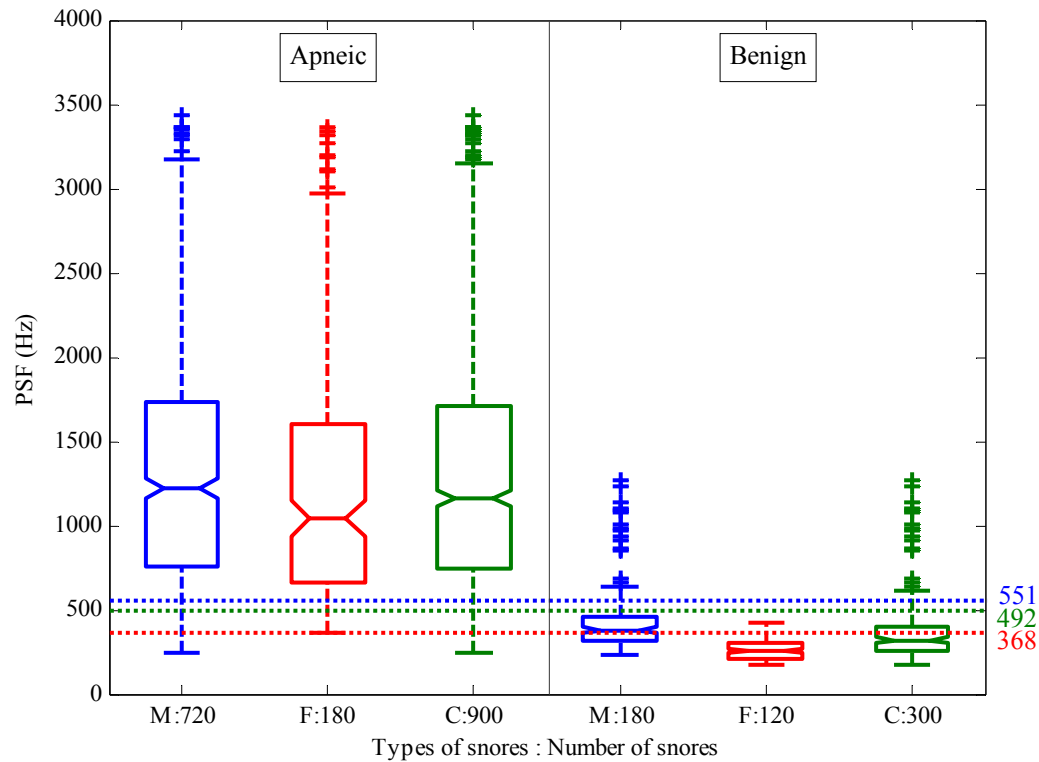
**Figure 6.7**    Notched box plots of peak sum frequency (PSF) of apneic and benign snores for males (M), females (F), and both males and females combined (C). Threshold values are marked on the right-hand side of the plots.

and PSF consistently yield superior sensitivity (PF1: 85.4–98.3%; PSF: 82.5–100%) and specificity (PF1: 85.0–100%; PSF: 83.3–100%). The classification results are graphically illustrated by means of ROC curves in Figure 6.5, in addition to notched box plots in Figures 6.6 and 6.7 for PF1 and PSF, respectively.

PF1 and PSF, which both signify approximated modes with highest coupling strength, have clearly demonstrated their capability to discriminate between AS and BS. To further appraise the diagnostic power of these two markers, we comparatively evaluated their performance with those obtained from the spectral PF. The computational procedures for PF are analogous to those mentioned in Section 5.6, except that the frame size has now changed to 512 samples, which is identical to the present experimental condition for WBC analysis, in order to conduct a fair test.

**Table 6.3**  Diagnostic performance of spectral peak frequency (PF) computed using the Welch's method (frame size = 512 samples or 46 ms) of apneic (A) and benign (B) snores for males (M), females (F), and both males and females combined (C).

| Type | AHI | Value (Hz) | Thre (Hz) | AUC | SE | Sens (%) | Spec (%) | p-value |
|------|-----|-----------|-----------|-----|-----|----------|----------|---------|
| M:A | 52.8 ± 24.7 | 604 ± 413 (95–2330) | 333 | 0.8163 | 0.0124 | 62.5 | 85.0 | < 0.0001 |
| M:B | 5.3 ± 3.5 | 248 ± 159 (88–867) | | | | | | |
| F:A | 23.6 ± 15.1 | 561 ± 554 (88–2612) | 213 | 0.9316 | 0.0069 | 91.7 | 97.5 | < 0.0001 |
| F:B | 3.4 ± 3.2 | 173 ± 100 (82–892) | | | | | | |
| C:A | 46.9 ± 25.7 | 596 ± 445 (88–2612) | 243 | 0.8537 | 0.0108 | 77.7 | 78.0 | < 0.0001 |
| C:B | 4.6 ± 3.4 | 218 ± 143 (82–892) | | | | | | |

Values are presented as mean ± standard deviation and range within brackets. AHI refers to apnea-hypopnea index in events/h; Thre, threshold in hertz; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

Results in Table 6.3 verify that the proposed markers can offer better diagnostic accuracy than PF. For both males and females combined, the threshold of PF = 243 Hz, with a resultant AUC = 0.8537, SE = 0.0108, sensitivity = 77.7%, specificity = 78.0%, and p-value < 0.0001. In all three patient groups, the thresholds of PF deliver a sensitivity of 62.5–91.7% and a specificity of 78.0–97.5%.

While both the parametric (see Chapter 5) and the nonlinear approaches achieve statistical significance between same- and both-gender patient groups (p-value < 0.0001), the latter approach yields a slightly higher specificity (83.3–100% versus 82.0–95.0%), along with a similar sensitivity of 82.5–100%, thereby suggesting that more benign patients can be correctly diagnosed as not having OSA.

## 6.5    Diagnostic Marker Regression Models

On top of the statistical measures, we attempted to best-fit the scattering data through various linear and nonlinear regression models [204-207], as pictured in Figures 6.8 and 6.9, in order to explain the feasible relationship between AHI and the proposed markers (PF1 and PSF). Under the same derivation procedures outlined in Section 5.7 for the regression equations, Table 6.4 and Table 6.5 list the forms of regression models for PF1 and PSF, respectively, along with the goodness-of-fit statistics ($R^2$ and RSD) and the predictive AHI values corresponding to the derived thresholds for both genders combined (PF1 = 285 Hz; PSF = 492 Hz).

Interestingly, although the quadratic functional form seems to fit the scattering data well ($R^2$ = 0.5555–0.5776), the dispersion of residuals between the predicted and the observed AHI values is far apart (RSD = 19.0152–19.5045 events/h) with broad confidence intervals, leading to inaccurate predictive AHI values (14.4–17.7 events/h). Ideally, the threshold of AHI should be 10.0 events/h as it was considered in this research work. In contrast to the quadratic model, the exponential and the power models perform better, in terms of RSD values (0.4347–0.5277 events/h) and AHI prediction (8.0–10.7 events/h), despite having lower $R^2$ (0.2911–0.5189). Since $R^2$ is not the major criterion for judging whether a fit is practical [204-207], one can deduce that the latter two regression models can realistically interpret the relationship between the severity of OSA and the potential dominant frequency modes involved in the nonlinear interactions in snore signals. The functional form of exponential or power are also in favor of illustrating the possible relationship between AHI and F1 of snore signals, as discussed in Section 5.7.
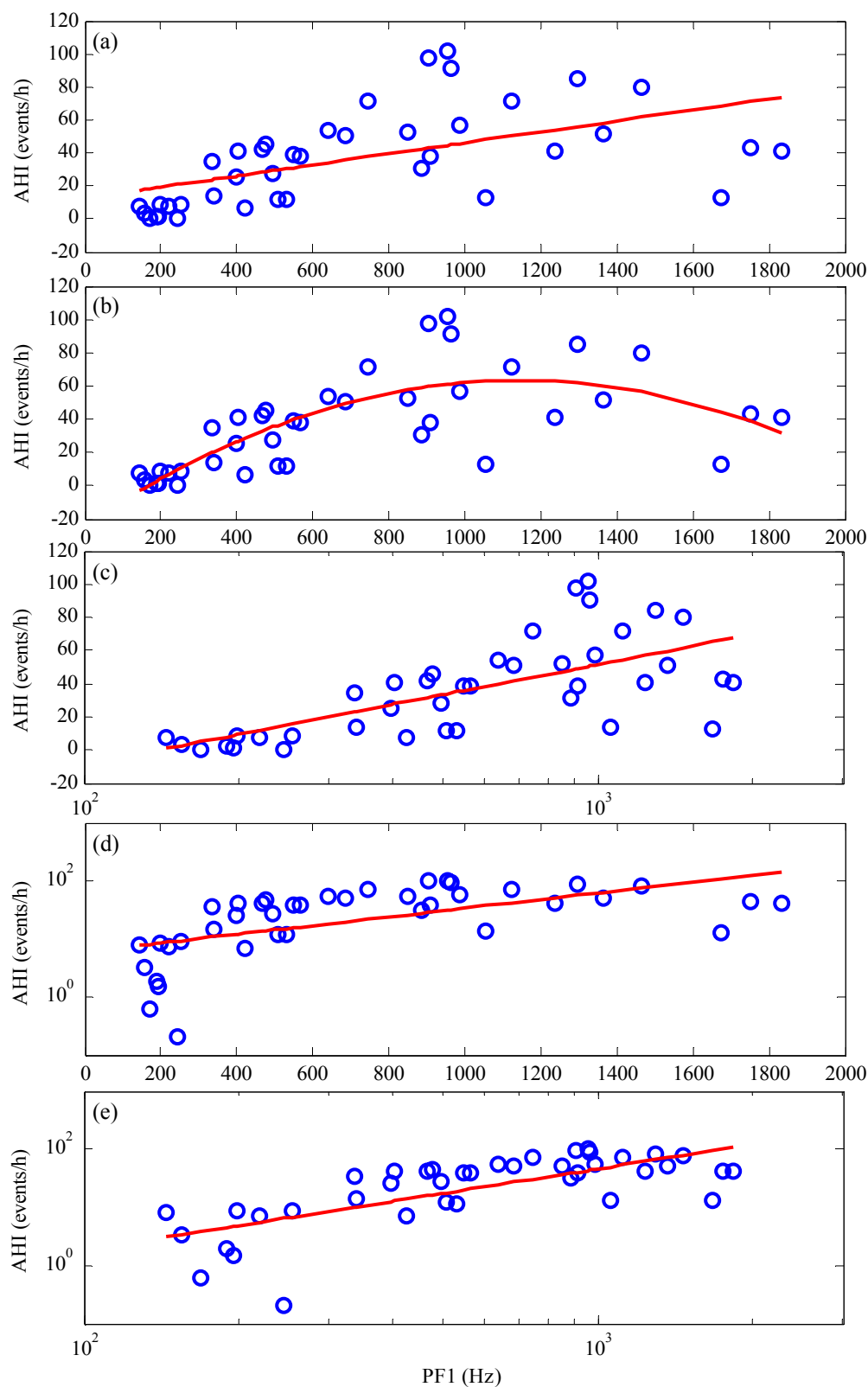
**Figure 6.8**    Regression models of (a) linear, (b) quadratic, (c) logarithmic, (d) exponential, and (e) power functions for apnea-hypopnea index (AHI) in events/h and peak frequency at $f_1$ (PF1) in hertz (Hz).
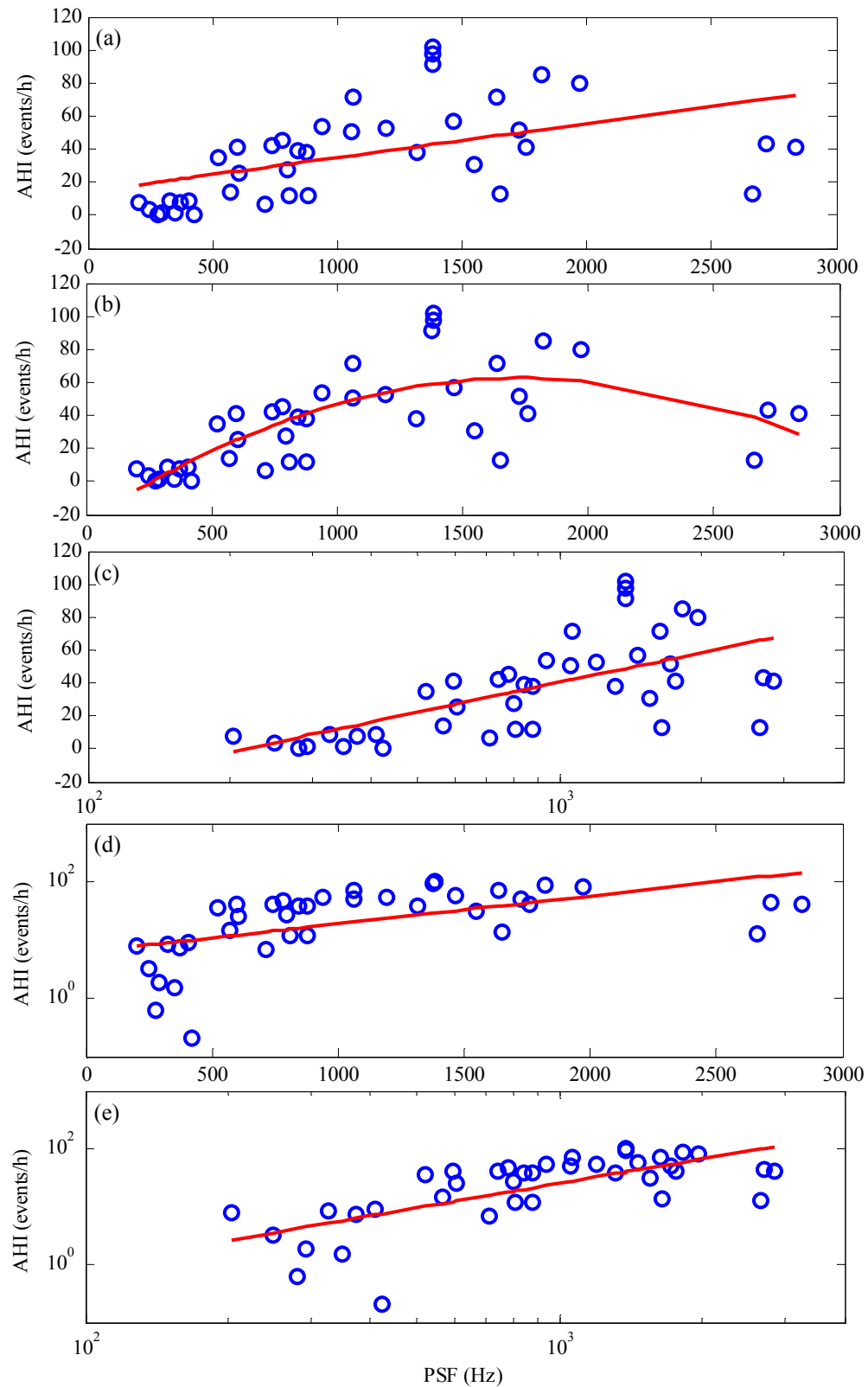
**Figure 6.9**    Regression models of (a) linear, (b) quadratic, (c) logarithmic, (d) exponential, and (e) power functions for apnea-hypopnea index (AHI) in events/h and peak sum frequency (PSF) in hertz (Hz).

**Table 6.4**   Regression models for apnea-hypopnea index (AHI) in events/h and peak frequency at $f_1$ (PF1) in hertz (Hz).

| Regression | Equation | $R^2$ | RSD | AHI[†] |
|---|---|---|---|---|
| Linear | AHI = 0.0335 (PF1) + 12.6229 | 0.3011 | 24.4586 | 22.2 |
| Quadratic | AHI = $-$ 0.0001 $(PF1)^2$ + 0.1512 (PF1) $-$ 23.3608 | 0.5776 | 19.0152 | 14.4 |
| Logarithmic | AHI = $-$ 129.4423 + 60.3602 Log (PF1) | 0.4463 | 21.7703 | 18.7 |
| Exponential | Log (AHI) = 0.7801  + 0.0007 (PF1) | 0.3269 | 0.5142 | 9.8 |
| Power | Log (AHI) = $-$ 2.5203 + 1.3943 Log (PF1) | 0.5189 | 0.4347 | 8.0 |

[†]Predictive AHI corresponding to derived thresholds for both males and females combined (PF1 = 285 Hz). Ideally, AHI = 10.0 events/h. $R^2$ refers to coefficient of determination; RSD, residual standard deviation in events/h; Log, logarithm function with base 10.

**Table 6.5**   Regression models for apnea-hypopnea index (AHI) in events/h and peak sum frequency (PSF) in hertz (Hz).

| Regression | Equation | $R^2$ | RSD | AHI[†] |
|---|---|---|---|---|
| Linear | AHI = 0.0207 (PSF) + 14.2461 | 0.2460 | 25.4041 | 24.4 |
| Quadratic | AHI = $-$ 0.00003 $(PSF)^2$ + 0.0996 (PSF) $-$ 24.3739 | 0.5555 | 19.5045 | 17.7 |
| Logarithmic | AHI = $-$ 139.7810 + 60.0108 Log (PSF) | 0.3995 | 22.6720 | 21.8 |
| Exponential | Log (AHI) = 0.7801  + 0.0007 (PF1) | 0.2911 | 0.5277 | 10.7 |
| Power | Log (AHI) = $-$ 2.8099 + 1.4036 Log (PSF) | 0.4761 | 0.4537 | 9.3 |

[†]Predictive AHI corresponding to derived thresholds for both males and females combined (PSF = 492 Hz). Ideally, AHI = 10.0 events/h. $R^2$ refers to coefficient of determination; RSD, residual standard deviation in events/h; Log, logarithm function with base 10.

## 6.6   Summary

This chapter provides entrances into the complex UA dynamics by examining the nonlinear properties of snore signals for the diagnosis of OSA. The analysis was performed by WBC and emphasized on the detection and quantification of QPC in nonstationary snore signals. Results reveal that nonlinear mode interactions in AS are less self-coupled and are usually associated with higher and wider frequency ranges than that of BS. Furthermore, the proposed diagnostic markers (PF1 and PSF) are indicative of AS and BS for all three patient groups (p-value < 0.0001), with a sensitivity of   82.5–100% and a specificity of 83.3–100%, outperforming the PF

whose sensitivity and specificity are in the range of 62.5–91.7% and 78.0–97.5%, respectively. These results corroborate our hypothesis that phase-coupled interaction modes in snore signals may facilitate the classification of snores generated by apneic and benign patients attributed to differences in their UA anatomical and functional abnormalities. Relationship between AHI and the proposed markers can possibly take the functional form of exponential or power, with a predictive AHI of 8.0–10.7 events/h.

Besides analyzing snore signal acoustically, the evaluation of snore sounds through perceptual means can also play a role in the diagnosis of OSA because strong correlations exist between acoustical and perceptual factors [232]. The next chapter of this thesis investigates the applicability of various psychoacoustic metrics to classify snores produced by apneic and benign patients.

Chapter 7

# Psychoacoustic Analysis and Classification of Snores

Snoring is foremost a subjective perception of the listener [25,183]. Prior to any kind of snore signal analysis, the signal per se must correspond to the sound that a listener perceives as a snore. Human hearing, therefore, remains the primary arbiter of what constitutes snoring. The hearing system is a complicated human organ, which constantly appraises the acoustic environment by means of sound perception mechanisms, also known as psychoacoustic metrics [232], for instance, loudness, sharpness, roughness, fluctuation strength, and even annoyance. Functionality of these metrics has been successfully realized in the analysis of auditory signals (e.g., speech [233-235], voice [236,237], and music [238,239]) and in noise control studies [240-244], ascertaining relationship between the acoustical and the perceptual space [232].

This chapter extends the application of psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) to snore sound analysis by exploring their capability to discriminate AS from BS. Formulations of the metrics, in terms of physical signal parameters, are outlined. Experimental conditions, comprising the dataset for evaluation, the listening test procedure, and the scale for rating sound quality, are also described. Results are quantitatively illustrated by ROC curves and notched box plots. Furthermore, we discussed the degree of association between the psychoacoustic metrics and the patient demographic and clinical data, such as body

_____

The contents of this chapter are partly based on the author's publications [A2,A7,A9].

mass index (BMI), neck circumference (NC), and AHI, using Pearson's product-moment and Spearman's rank correlation coefficients.

## 7.1    Psychoacoustic Metrics

### 7.1.1    Loudness

Loudness is an essential parameter for any sound quality survey. It is a measure of sound strength defined by [232]

$$N = \int_0^{24\,\text{Bark}} N' dz \qquad (7.1)$$

where $N$ is the loudness in sone that is referenced to a 1 kHz tone at a sound pressure level of 40 dB, $N'$ is the specific loudness in sone/Bark, and $dz$ is the increment in the critical band rate $z$. The specific loudness represents the amount of loudness attributed to the auditory filters, and it is closely related to the hearing system excitation $E$ through

$$N' = 0.08 \left( \frac{E_{TQ}}{E_0} \right)^{0.23} \left[ \left( 0.5 + 0.5 \frac{E}{E_{TQ}} \right)^{0.23} - 1 \right] \qquad (7.2)$$

where $E_{TQ}$ is the excitation at threshold in silence, while $E_0$ is the excitation corresponding to the reference intensity $I_0 = 10^{-12}$ W/m$^2$. The specific loudness serves as a basis for many other psychoacoustic metrics.

### 7.1.2    Sharpness

Sharpness is an attribute that interprets the tone color of a sound, in terms of its powerfulness or aggressiveness. It relies on the weighted centroid of the specific loudness content, as highlighted in [232]

$$S = 0.11\frac{\int_0^{24\,\text{Bark}} N' g(z) z \, dz}{\int_0^{24\,\text{Bark}} N' \, dz} \tag{7.3}$$

where $S$ denotes the sharpness in acum, and $g(z)$ denotes the weighing function of the critical band rate $z$. An acum is referenced to a band of noise centered at 1 kHz with 60 dB. A higher sharpness value specifies a greater energy in high frequencies.

### 7.1.3    Roughness

Roughness is a sensation arising from rapid (15–300 Hz) temporal variations of a sound caused by beats between tones in a critical band or by amplitude- or frequency-modulated tones. A 1 kHz tone at 60 dB with 100% amplitude modulation at 70 Hz can create the roughness $R$ of 1 asper [232]. Explicitly,

$$R = 0.3 \frac{f_{\text{mod}}}{\text{kHz}} \int_0^{24\,\text{Bark}} \frac{\Delta L_E(z) dz}{\text{dB/Bark}} \tag{7.4}$$

where $f_{\text{mod}}$ signifies the modulation frequency, and $\Delta L_E$ signifies the range of excitation level within an auditory filter. Roughness increases with increasing modulation depth of the sound temporal masking pattern.

### 7.1.4    Fluctuation Strength

Fluctuation strength reveals human sensitivity towards slow (up to 20 Hz) moving amplitude modulation for a sound with frequency modulated at approximately 4 Hz. The unit of fluctuation strength is vacil, referenced to a 1 kHz tone at 60 dB with 100% amplitude modulation at 4 Hz, and the fluctuation strength of a sound can be expressed as [232]

$$F = \frac{0.008 \int_0^{24\,\text{Bark}} (\Delta L/\text{dB Bark}) \, dz}{(f_{\text{mod}}/4\text{Hz}) + (4\text{Hz}/f_{\text{mod}})} \tag{7.5}$$

where $\Delta L$ is the masking depth or the difference between the maxima and the minima in the temporal masking pattern. Fluctuation strength is associated with the fluent speech at a speaking rate of 4 syllables/s because its amplitude modulation is concentrated around 4 Hz.

### 7.1.5     Annoyance

Annoyance $PA$ is a mixture of different hearing sensations including loudness $N$, sharpness $S$, roughness $R$, and fluctuation strength $F$. It can be estimated by [232]

$$PA \approx N_5\left(1+\sqrt{w_s^2 + w_{FR}^2}\right) \tag{7.6}$$

where

$$w_s = (S-1.75)0.25\log_e(N_5+10), \qquad S>1.75 \tag{7.7}$$

and $$w_{FR} = \frac{2.18}{(N_5)^{0.4}}(0.4F+0.6R) \tag{7.8}$$

with $N_5$ indicating the percentile loudness in sone. It is apparent from Equation (7.6) that loudness has a dominant impact in the estimation of annoyance.

## 7.2     Experimental Conditions

### 7.2.1     Patient Dataset

Among the 40 patients listed in Table 1.1, 13 patients (8 apneic and 5 benign) participated in this psychoacoustic research. Table 7.1 contains the patient demographic and clinical data. Similar to those mentioned in Section 5.5.1, snoring sounds were recorded through the acquisition system for snore signals (see Chapter 3), and then preprocessed via the wavelet-based preprocessing system (see Chapter 4). A snore sound sample, which consists of five consecutive snoring episodes, lasting about 30 s, was selected from each patient, giving a total of 13 samples for evaluation.

**Table 7.1**  Demographic and clinical data of patients participated in psychoacoustic research.

| Type | Sample size | Age (years) | BMI (kg/m$^2$) | NC (cm) | AHI (events/h) |
|---|---|---|---|---|---|
| Combined : Apneic | 8 (M:6 ; F:2) | 42 ± 13 (23–59) | 34.9 ± 8.4 (25.0–50.4) | 42.5 ± 4.5 (36.0–51.0) | 55.0 ± 25.6 (14.2–90.8) |
| Combined : Benign | 5 (M:4 ; F:1) | 42 ± 10 (28–52) | 24.3 ± 2.9 (21.6–28.1) | 38.5 ± 2.0 (35.0–40.0) | 4.0 ± 3.5 (0.2–8.6) |

Values are presented as mean ± standard deviation and range within brackets. BMI refers to body mass index in kg/m$^2$; NC, neck circumference in centimeters; AHI, apnea-hypopnea index in events/h; M, males; F, females; Combined, both males and females combined.
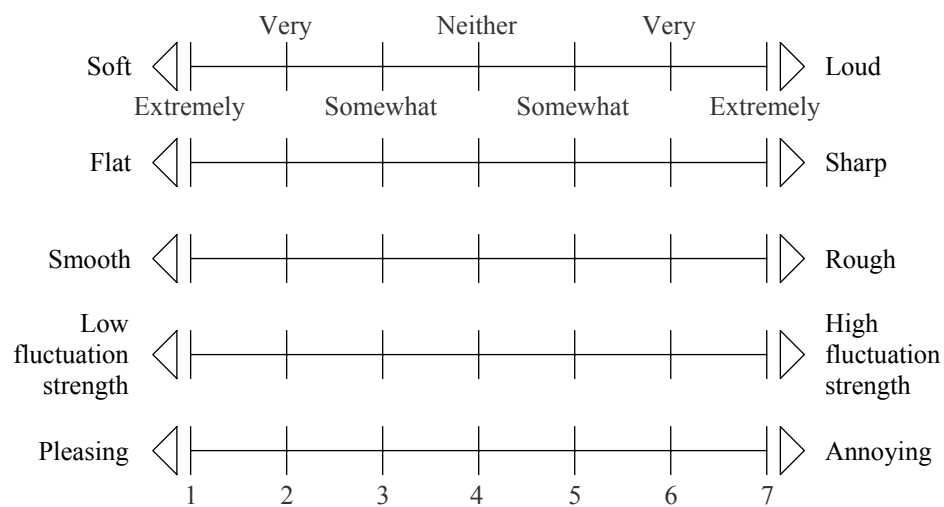


**Figure 7.1**   A 7-point semantic differential scale with bipolar adjective pairs.

## 7.2.2    Listening Test Procedure

A snore sound sample was presented three times continuously in a quiet room to 25 normal-hearing listeners (8 polysomnographic technicians from the Sleep Disorders Unit and 17 signal processing specialists from our institution), one at a time. Each listener was asked to pass judgments on the sample based on the five psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) via a 7-point semantic differential scale (range 1–7) with bipolar adjective pairs [232,245], as depicted in Figure 7.1, before continuing to the next sample. We chose the semantic differential scale in this research work owing to its aptitude to valuably convey the

**Table 7.2**   Diagnostic performances of psychoacoustic metrics of apneic and benign snore sounds.

| Metric | Thre | AUC | SE | Sens (%) | Spec (%) | p-value |
|---|---|---|---|---|---|---|
| Loudness | 4 | 0.9328 | 0.0136 | 78.0 | 91.2 | < 0.0001 |
| Sharpness | 4 | 0.8500 | 0.0207 | 60.5 | 88.0 | < 0.0001 |
| Roughness | 4 | 0.8978 | 0.0169 | 74.0 | 81.6 | < 0.0001 |
| Fluctuation strength | 4 | 0.8392 | 0.0214 | 63.0 | 81.6 | < 0.0001 |
| Annoyance | 5 | 0.9265 | 0.0142 | 72.0 | 92.0 | < 0.0001 |

Thre refers to threshold; AUC, area under receiver operating characteristic curve; SE, standard error of AUC; Sens, sensitivity in percentage; Spec, specificity in percentage; p-value, value of statistical significance which was considered to be present when p-value < 0.05.

denotative (e.g., loudness and sharpness) and the connotative (e.g., annoyance) meanings of a sound [232,245].

## 7.3     Diagnostic Results and Discussion

### 7.3.1     Statistical and Exploratory Data Analysis

Based on the dataset of 13 snore sound samples and sample size of 25 listeners, the diagnostic accuracy was quantified using ROC curves [155-157] and notched box plots [139,140,181,182] whose respective algorithms are explained in Section 4.3.4 and Section 5.5.1. However, due to the limited sound samples, the diagnostic sensitivity and specificity were determined from the same dataset; both training and test data are identical. A p-value < 0.05 was considered statistical significant. Results from the ROC curve analysis are summarized in Table 7.2, accompanied by ROC curves and notched box plots in Figures 7.2 and 7.3, respectively.

Apneic snore sounds achieve higher ratings for all the metrics than the benign ones, implying that the qualities of apneic snore sounds are louder, sharper, rougher, higher fluctuation strength, and more annoying. Loudness and annoyance rank the top two metrics, in sequence, that best distinguish AS from BS (AUC = 0.9265–0.9328, SE = 0.0136–0.0142, sensitivity = 72.0–78.0%, specificity = 91.2–92.0%, and p-value
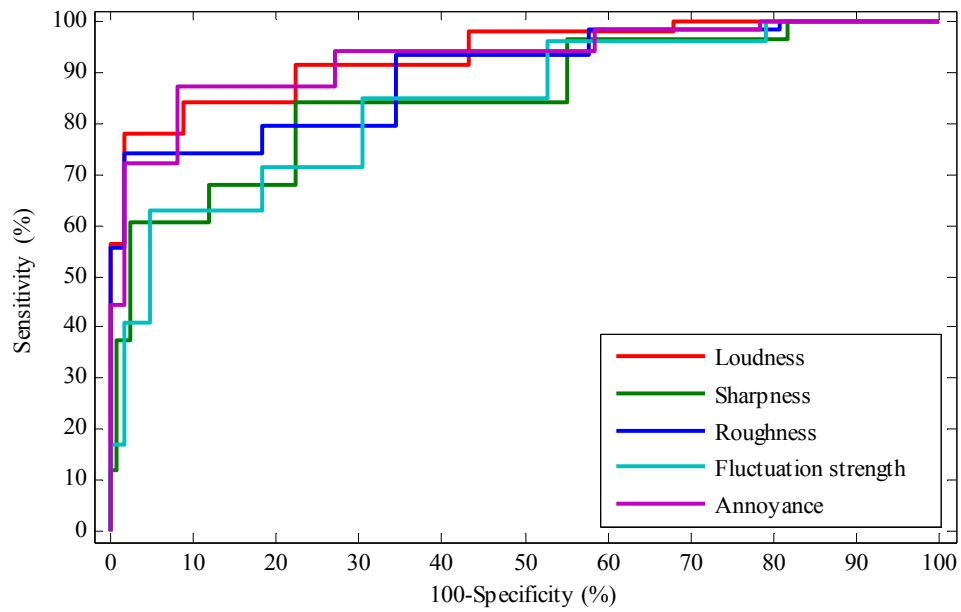
**Figure 7.2**    Receiver operating characteristic curves of psychoacoustic metrics of apneic and benign snore sounds.
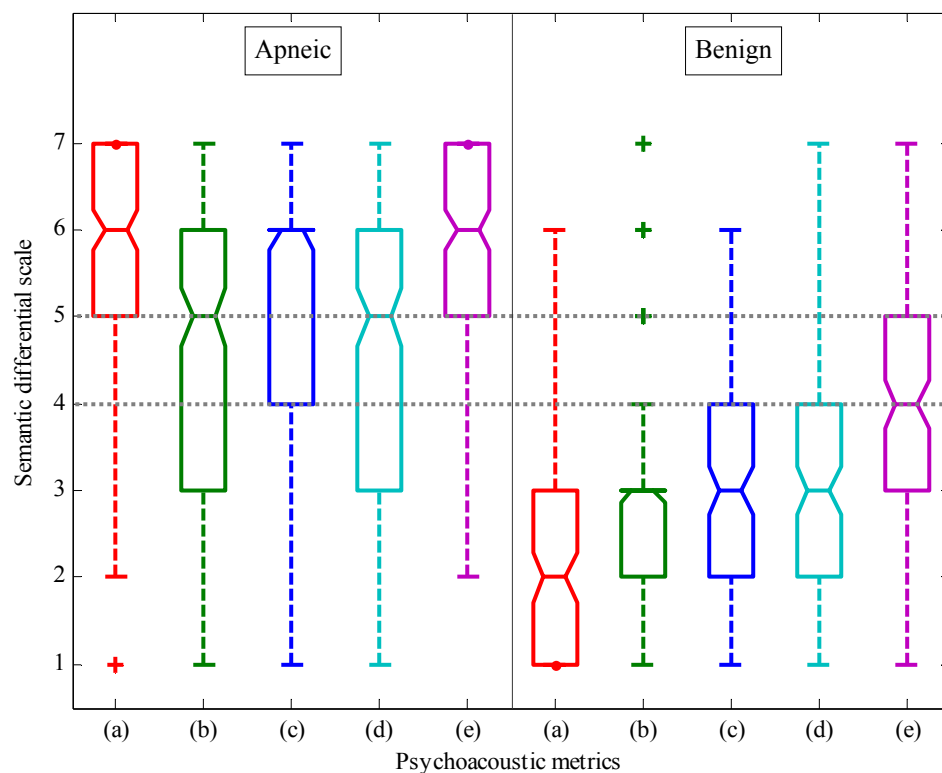


**Figure 7.3**    Notched box plots of psychoacoustic metrics: (a) loudness; (b) sharpness; (c) roughness; (d) fluctuation strength; and (e) annoyance of apneic and benign snore sounds. Threshold values are marked.

< 0.0001), followed by roughness, sharpness, and fluctuation strength (AUC = 0.8392–0.8500, SE = 0.0169–0.0214, sensitivity = 60.5–74.0%, specificity = 81.6–88.0%, and p-value < 0.0001).

Moreover, the typical threshold separating these two diagnostic classes of snore sounds falls on the rating of 4 (neutral response), except for the metric of annoyance whose threshold is 5, emphasizing that snoring is often considered as a nuisance. A study on 37 consecutive snoring men has found that 55% of their bed partners are irritated by snoring, and 40% of them chose to sleep in a separate room more than once a week in order to get a more restful and recuperative sleep [246], negatively affecting their partners' health and well-being [105,247,248]. Other studies have further reported the devastating consequences of snore-induced noise pollution on couple's relationship, for example, marital strife, divorces, and attempted murder [249-251].

### 7.3.2    Correlation Analysis

To examine the association between the psychoacoustic metrics and the patient demographic and clinical data (BMI, NC, and AHI), we employed the Pearson's product-moment and the Spearman's rank correlation measures [139,140]; the former correlation coefficient $r_p$ valuates the strength of a linear relationship between two variables, whereas the latter correlation coefficient $r_s$ is a non-parametric measure of monotone association between a pair of variables, as pointed out in Section 4.1.3. Nonetheless, both the coefficients lie in the range of -1 and +1, where the negative or positive sign indicates a negative or positive association, respectively, along with the magnitude that reflects the association strength. A correlation coefficient of zero means that there is entirely no association between the two variables; both variables

**Table 7.3**     Pearson's product-moment and Spearman's rank correlation results between psychoacoustic metrics and patient demographic and clinical data.

| Metric | BMI (kg/m$^2$) | | NC (cm) | | AHI (events/h) | |
|---|---|---|---|---|---|---|
| | $r_p$ | $r_s$ | $r_p$ | $r_s$ | $r_p$ | $r_s$ |
| Loudness | 0.4963 | 0.5878 | 0.5564 | 0.6828 | 0.7432 | 0.7062 |
| Sharpness | 0.3010 | 0.3687 | 0.4052 | 0.4739 | 0.5668 | 0.5112 |
| Roughness | 0.3687 | 0.4701 | 0.4867 | 0.5879 | 0.6835 | 0.6520 |
| Fluctuation strength | 0.2885 | 0.3400 | 0.4795 | 0.5069 | 0.5796 | 0.5304 |
| Annoyance | 0.4356 | 0.5340 | 0.5313 | 0.6638 | 0.7182 | 0.7162 |

BMI refers to body mass index in kg/m$^2$; NC, neck circumference in centimeters; AHI, apnea-hypopnea index in events/h; $r_p$, Pearson's product-moment correlation coefficient; $r_s$, Spearman's rank correlation coefficient. All p-values are < 0.0001, indicating statistically significant correlations.

move independently.

Results from the parametric and non-parametric correlation analysis in Table 7.3 evidently demonstrate that loudness and annoyance are the two most preferable psychoacoustic metrics for assessing snore sounds, with loudness being the best metric as it yields the greatest association strength with AHI ($r_p$ = 0.7432; $r_s$ = 0.7062), NC ($r_p$ = 0.5564; $r_s$ = 0.6828), and BMI ($r_p$ = 0.4963; $r_s$ = 0.5878). Conversely, roughness, sharpness, and fluctuation strength are less dependent on AHI ($r_p$ = 0.5668–0.6835; $r_s$ = 0.5112–0.6520), NC ($r_p$ = 0.4052–0.4867; $r_s$ = 0.4739–0.5879), and BMI ($r_p$ = 0.2885–0.3687; $r_s$ = 0.3400–0.4701). These outcomes are in conformity with those obtained from the ROC curve analysis. All p-values computed from the correlation analysis are less than 0.0001, indicating statistically significance.

A close inspection of the correlation results also suggest that all the psychoacoustic metrics are more closely related to the NC than the BMI, while both NC and BMI are good predictors of AHI [252-254]. A likely rationalization is that the neck measurements correspond more directly to the UA soft tissues that possess an immediate influence on the pathophysiology of snoring and OSA [6,7,25]. Excessive

soft tissues surrounding the UA, which is in the case of most apneic patients [29-31,56,57], considerably reduce the airway space and vibrate vigorously when a constant airflow passes through the constrictions in the UA, thereby increasing the complexity of acoustic source dynamics (e.g., frequency or harmonic pattern and temporal regularity) and producing amplified sound perception. Correspondingly, snore sounds generated by apneic patients are unquestionably more annoying than those generated by benign patients since annoyance is a composite psychoacoustic metric with loudness as the key contributor.

## 7.4     Summary

This chapter explores the applicability of psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance) in the classification of AS and BS. Snore sound samples from 13 patients were appraised by 25 listeners by means of a 7-point semantic differential scale with bipolar adjective pairs. Results from various statistical and correlation analysis consistently show differences between the psychoacoustics of apneic and benign snore sounds (p-value < 0.0001). Among the five psychoacoustic metrics, loudness and annoyance are the two most potential diagnostic markers for OSA, in sequence, because they can deliver a higher diagnostic accuracy (sensitivity = 72.0–78.0%, specificity = 91.2–92.0%) and stronger association with AHI ($r_p$ = 0.7182–0.7432; $r_s$ = 0.7062–0.7162), NC ($r_p$ = 0.5313–0.5564; $r_s$ = 0.6638–0.6828), and BMI ($r_p$ = 0.4356–0.4963; $r_s$ = 0.5340–0.5878), as compared to roughness, sharpness, and fluctuation strength.

Several promising snore-based OSA markers have been proposed in this thesis with remarkable diagnostic results. To provide more concrete justifications on the validity of the proposed markers, the next chapter of this thesis attempts to evaluate

the relationships between the snore source flow, the UA anatomical structures, and the

properties of snores.

Chapter 8

# Snore Physiological-Anatomical-Acoustical Relationships

While snore-based diagnostic markers appear to shed light on the detection of OSA both in the literature (see Chapter 2) and in this thesis, there is little research on the correlations between the snore source flow (SF, i.e., acoustic source in snore production), the UA anatomical structures (e.g., cross-sectional airway dimensions), and the characteristics of snores. Such correlation studies are undeniably crucial because they not only provide insights into different perspectives on snores from physiological to acoustical and perceptual domains, but also warrant the feasibility of using snore signals as an alternative to PSG in diagnosing OSA.

This chapter attempts to address this concern through the following two studies. First, it investigates the acoustical and perceptual influences of changing the cross-sectional areas (CSA) of the pharynx (PX) and oral cavity (OC) on the generation of snores, which correspondingly validates the reliability of the proposed formant frequency (F1), spectral peak frequency (PF), and psychoacoustic metrics (loudness, sharpness, roughness, fluctuation strength, and annoyance). Second, it parameterizes and models the SF and its derivative (SFD), in addition to generating synthetic snores, with a hypothesize that temporal and spectral attributes of SF and SFD may contain essential information about snore excitation source (ES) dynamics and affect the characteristics of snores, both acoustically and perceptually.

_____

The contents of this chapter are partly based on the author's publications [A1,A2].

To accomplish the study tasks, we estimated SFs directly from natural snores, developed acoustic models of the UA, and synthesized snores by perturbing the CSA of PX and OC, as well as by utilizing various SFD pulse shapes. Methodologies for SF analysis (signal estimation, parameterization, and modeling), UA acoustic modeling, and snore synthesis are described in detail with graphical illustrations, followed by experimental results, in terms of objective and subjective measures.

## 8.1     Snore Source Flow Estimation

The soft tissue in the UA is a flow-induced self-excited biomechanical oscillator [208,255-257]; it acts as the main ES during the production of snores. Direct measuring of airflow at snore ES, defined as SF, is a complicated process attributable to its difficulty of accessing the snoring site during natural (without anesthesia) nocturnal snoring. Nevertheless, if one applies the well-known source-filter theory [152,158-161], as discussed in Section 5.1, in snore generation by presuming that the ES and the UA are linearly separable, and that the SF is acoustically filtered by the UA and the lip radiation to produce snoring sounds, a feasible technique for estimating SF is the inverse filtering – a technique extensively applied in speech science to explore the nature of vocal fold excitation, indirectly and non-invasively [152,158-161,258].

Figure 8.1 depicts a flow chart of an iterative adaptive inverse filtering technique [258] for approximating SF. To offer a clearer picture of the technique, graphical illustrations of a quasi-periodic snore signal processing through the technique are shown alongside the figure. This technique removes the effects of ES and UA from a snore signal through an iterative structure that repeats twice, with a goal to acquire more accurate SF estimate as compared to a direct inverse filtering method. The technique robustness has been demonstrated in voice source analysis for the estimation
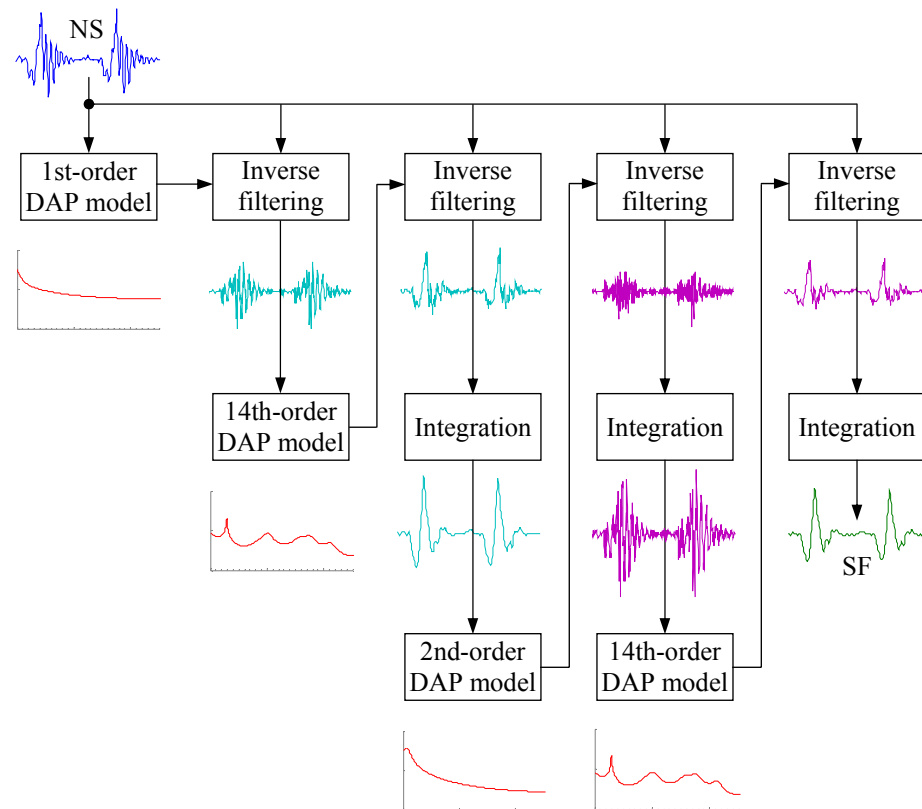
**Figure 8.1**    Flowchart of an iterative adaptive inverse filtering technique with discrete-all pole (DAP) modeling. NS and SF denote natural snore and source flow, respectively.

of glottal flow [258-260], and it consists of the following two phases. First, an initial estimate of SF is computed from a denoised snore by eliminating the contributions of ES, UA, and lip radiation, correspondingly simulated by a 1st-order and a 14th-order DAP model, and a fixed differentiator. The 1st-order DAP model approximates the combined effect (-6 dB/octave) of the ES (-12 dB/octave) and the lip radiation (+6 dB/octave), and it is the first to remove from the input signal through inverse filtering. Subsequently, the effect of UA, represented by a 14th-order DAP model, is canceled from the input signal by inverse filtering prior to eliminating the lip radiation effect via integration (i.e., the resulting signal is filtered with the inverse of the differentiator). This concludes the first phase of inverse filtering and yields the initial SF estimate. Second phase of the technique begins with a 2nd-order DAP modeling of the initial SF

estimate to account for the approximated contribution of ES, which is removed from the input signal through inverse filtering, followed by an integration of its output signal to eliminate the lip radiation effect. Eventually, the final SF estimate is determined after canceling the contributions of UA and lip radiation by inverse filtering the input signal with a 14th-order DAP model obtained in the second phase and by integrating the resulting signal, respectively.

To give simple yet reasonable approximations for the filter coefficients, we implemented the DAP modeling [173] in the iterative adaptive inverse filtering technique rather than the LP modeling whose spectral peaks are highly biased towards the pitch harmonics, as elaborated in Section 5.2. Moreover, we evaluated the order of the DAP model representing the UA, apart from using Equations (5.44)–(5.47), through several criteria [160,261] including the final prediction error [262]

$$\text{FPE}(p) = \sigma^2 \left( \frac{N + (p+1)}{N - (p+1)} \right),$$
(8.1)

the Akaike information criterion [263]

$$\text{AIC}(p) = N \log_e (\sigma^2) + 2p,$$
(8.2)

and the minimum description length [264]

$$\text{MDL}(p) = N \log_e (\sigma^2) + p \log_e (N).$$
(8.3)

These criteria generally weight the prediction error variance $\sigma^2$, accompanied by the frame size $N$, and determine the model order that delivers the minimum error value.

Figure 8.2 exhibits the three typical forms of snore signals encountered in the patient dataset detailed in Section 5.5.1 and in the literature [28,97,100,216]. These signals can be mathematically formulated [100] and broadly characterized as quasi-periodic or aperiodic. The two quasi-periodic snores, namely snore 1 and snore 2, were produced by two male patients whose AHI values are 45.6 events/h and 7.0 events/h,
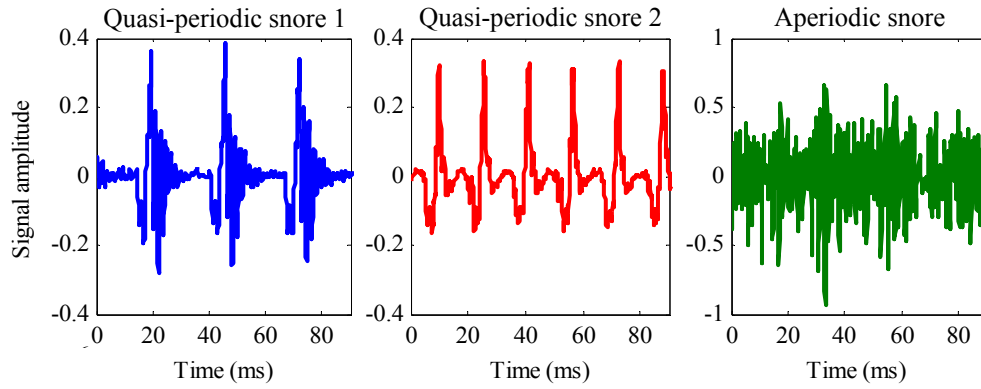
**Figure 8.2**  Typical waveforms of snores.
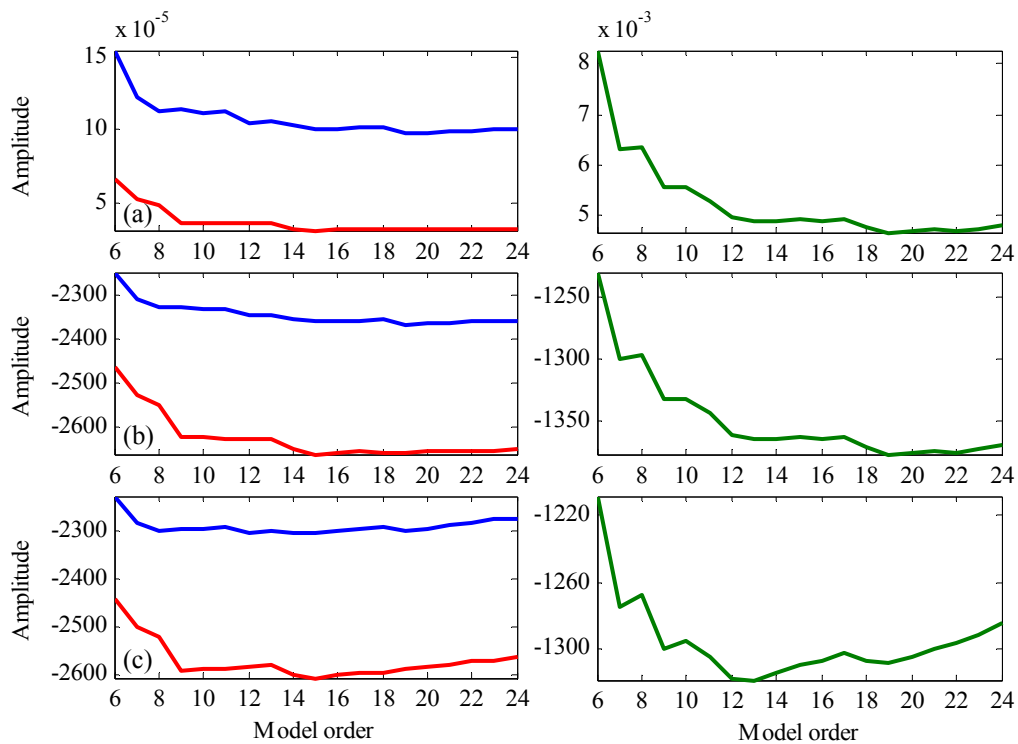


**Figure 8.3**  Performance of different model order selection criteria: (a) final prediction error; (b) Akaike information criterion; (c) minimum description length for quasi-periodic snore 1 (blue) and snore 2 (red), as well as aperiodic snore (green).

respectively; on the other hand, the aperiodic snore was produced by a male patient whose AHI value is 41.1 events/h.

The criterion values of the three snores at $N = 256$ samples ($\approx$ 23 ms) are displayed in Figure 8.3. As can be noted, the derived $p = 14$ in Section 5.5.2 remains a
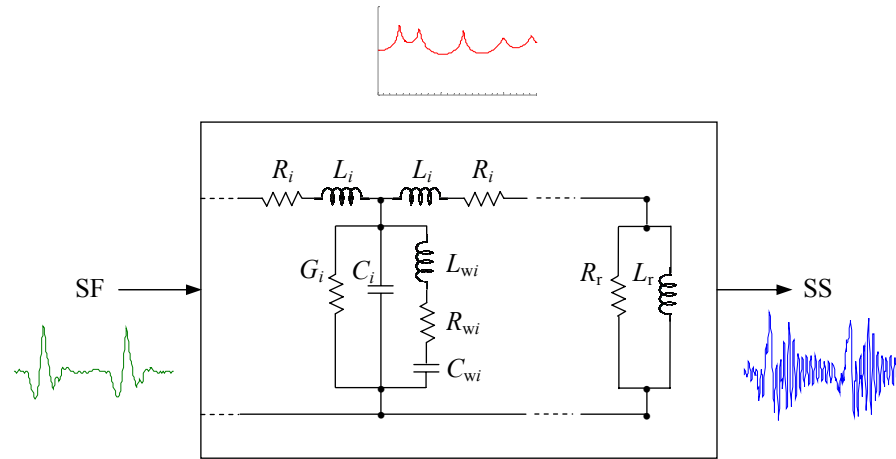
**Figure 8.4**    Transmission-line model of the upper airway for $i$th section and lip radiation load with lumped parameters including resistance $R_i$, inertance $L_i$, compliance $C_i$, conductance $G_i$, wall inertance $L_{wi}$, wall resistance $R_{wi}$, wall compliance $C_{wi}$, radiation resistance $R_r$, and radiation inertance $L_r$. SF and SS denote source flow and synthetic snore, respectively.

preferable choice to model the UA attributed to its relatively low error values for all the criteria, further justifying the use of a 14th-order DAP model to configure the UA.

## 8.2    Upper Airway Acoustic Modeling

Since both snore and speech sounds propagate through the same medium, an UA model that may ascertain the anatomical-acoustical relationships is likely comparable to the transmission-line model of the lossy vocal tract with non-rigid walls, where wave propagation in an acoustic tube resembles plane wave propagation along an electrical transmission line  [152,158-161,265,266]. A transmission-line model of the UA is rendered in Figure 8.4; alongside the figure are graphically illustrations of snore synthesis. In spite of the UA inherent intricacy, it can be regarded as a concatenation of cylindrical sections whose lengths are shorter than 1/8th the acoustic wavelength of interest [158,159,266]. With the maximum frequency of interest for snore signals as 5000 Hz [99,105] and the speed of sound for moist air at 37ºC as 35400 cm/s, the

**Table 8.1**   Lumped parameters for $i$th section of upper airway model and lip radiation load.

| Parameter | Expression | Parameter | Expression |
|---|---|---|---|
| Resistance | $R_i = \dfrac{l_i S_i}{2 A_i^2} \sqrt{\dfrac{\rho \mu \omega}{2}}$ | Wall resistance | $R_{wi} = \dfrac{b}{l_i S_i}$ |
| Inertance | $L_i = \dfrac{\rho l_i}{2 A_i}$ | Wall inertance | $L_{wi} = \dfrac{m}{l_i S_i}$ |
| Compliance | $C_i = \dfrac{l_i A_i}{\rho c^2}$ | Wall compliance | $C_{wi} = \dfrac{l_i S_i}{k}$ |
| Conductance | $G_i = l_i S_i \dfrac{\eta - 1}{\rho c^2} \sqrt{\dfrac{\lambda \omega}{2 \xi \rho}}$ | Radiation resistance | $R_r = \dfrac{128 \rho c}{9 \pi^2 A_m}$ |
| | | Radiation inertance | $L_r = \dfrac{8 \rho}{3 \pi \sqrt{\pi A_m}}$ |

$l_i$, $S_i$, and $A_i$ refer to length, circumference, and cross-sectional area of the $i$th cylindrical section, respectively; $\omega$, radian frequency; $\rho$, density of air = $1.14 \cdot 10^{-3}$ g·cm$^{-3}$ (moist air 37°C); $c$, sound velocity = $3.54 \cdot 10^4$ cm·s$^{-1}$ (moist air 37°C); $\mu$, shear viscosity = $1.86 \cdot 10^{-4}$ dyne·s·cm$^{-2}$ (20°C, 1 atm); $\lambda$, heat conduction coefficient of air = $5.5 \cdot 10^{-5}$ cal·cm$^{-1}$·s$^{-1}$·°C$^{-1}$ (0°C); $\eta$, adiabatic gas constant = 1.4; $\xi$, specific heat of air at constant pressure = 0.24·cal·g$^{-1}$·°C$^{-1}$ (0°C, 1 atm); $m$, $b$, and $k$ are the mass = 1.5 g, mechanical resistance = $1.6 \cdot 10^3$ g·s$^{-1}$, and stiffness = $3 \cdot 10^4$ dyne·cm$^{-1}$ of wall per unit area, respectively; $A_m$, cross-sectional area of mouth, which was deemed as the area of the last section.

maximum section length is found to be 0.89 cm. In this research work, we adopted an equal section length of 0.10 cm, which increases the number of sections representing the UA, thereby improving the accuracy of the UA acoustic modeling.

Losses owing to viscous friction, heat conduction, and tissue vibration in each section are represented by frequency-dependent lumped parameters, as listed in Table 8.1, together with the physical properties of air. The acoustic resistance $R_i$ arises from viscous and thermal losses at the boundaries, and the conductance $G_i$ from heat conduction on the $i$th section walls. The inertance $L_i$ is associated with the mass of air, and the compliance $C_i$ is with the ability of air to expand and compress. The series combination of mechanical wall impedance comprising a resistance $R_{wi}$, inertance $L_{wi}$, and compliance $C_{wi}$ model the effect of wall vibration. Furthermore, radiation load at the lips is configured by a parallel connection of resistance $R_r$ and inertance $L_r$ to

account for energy loss and mass inertia of air. The UA acoustic transfer function can be ultimately described by the product of all section transfer matrices.

As an explanatory example, consider the UA with $N$ cylindrical sections. The UA acoustic transfer matrix

$$T(w) = T_1(w)T_2(w)\ldots T_{N-1}(w)T_N(w) \tag{8.4}$$

where $w$ is the angular frequency,

$$T_i(w) = \begin{bmatrix} 1 + \dfrac{Z_{ai}}{Z_{bi}} & 2Z_{ai} + \dfrac{Z_{ai}^2}{Z_{bi}} \\ \dfrac{1}{Z_{bi}} & 1 + \dfrac{Z_{ai}}{Z_{bi}} \end{bmatrix} \tag{8.5}$$

is the $i$th section transfer matrix derived from impedances

$$Z_{ai} = Z_{0i} \tanh\left(\frac{\gamma_i \ell_i}{2}\right) \tag{8.6}$$

and

$$Z_{bi} = \frac{Z_{0i}}{\sinh(\gamma_i \ell_i)}. \tag{8.7}$$

These impedances are correlated with the section length $\ell_i$, characteristic impedance, and propagation constant [152,158-161], where the latter two variables are respectively equated as

$$Z_{0i} = \sqrt{\frac{z_i}{y_i}} \tag{8.8}$$

and

$$\gamma_i = \sqrt{z_i y_i} \tag{8.9}$$

with

$$z_i = R_i + jwL_i \tag{8.10}$$

and

$$y_i = G_i + jwC_i + \left(\frac{1}{R_{wi} + jwL_{wi} + (1/jwC_{wi})}\right) \tag{8.11}$$

describing the impedance of medium and the admittance of medium and wall, correspondingly.
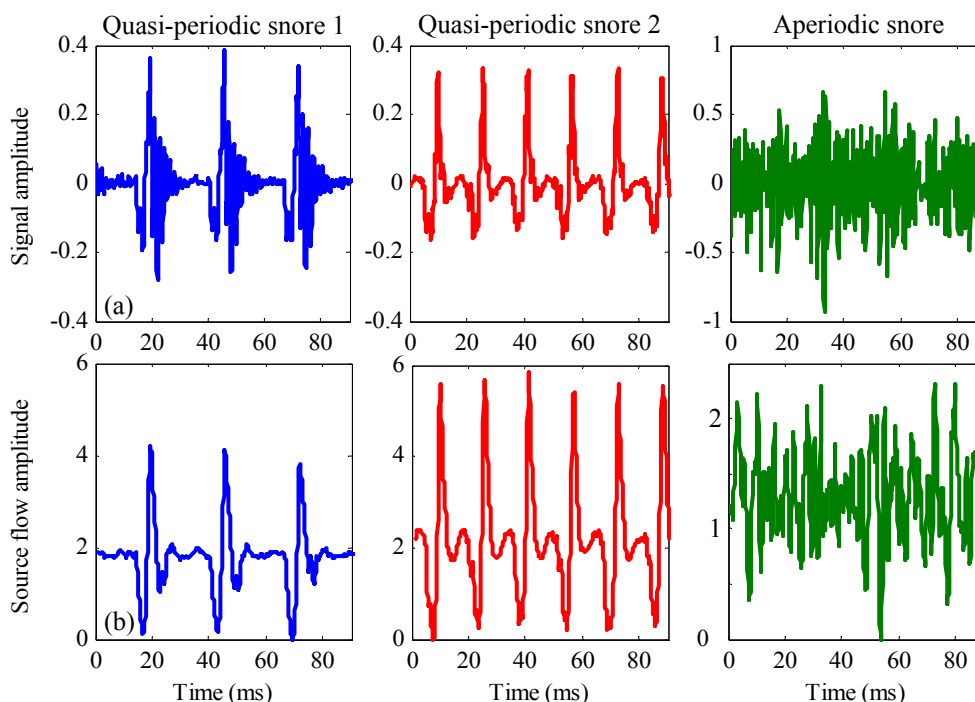
**Figure 8.5**    Waveforms of (a) quasi-periodic and aperiodic snores, and their corresponding (b) source flow waveforms.

## 8.3    Synthetic Snore Generation

### 8.3.1    Experimental Conditions for Area Perturbation Study

#### 8.3.1.1    Patient Dataset

The iterative adaptive inverse filtering technique [258] with DAP modeling [173] was applied on the three typical forms of snore signals in Figure 8.2, and the outcomes are shown in Figure 8.5. The two quasi-periodic signals yield rhythmically repeating SF waveforms. Strong peaks in the waveforms reveal the periodic nature of the rate at which the ES maneuvers (fundamental frequency). The first quasi-periodic SF, designated as SF1, has a longer plateau of airflow and a lower fundamental frequency (38 Hz) than the other quasi-periodic SF designated as SF2  (64 Hz). Conversely, the SF of the aperiodic snore signal, denoted as SF3, is random with high-frequency chaotic oscillations and no fundamental frequency. The fundamental frequencies were
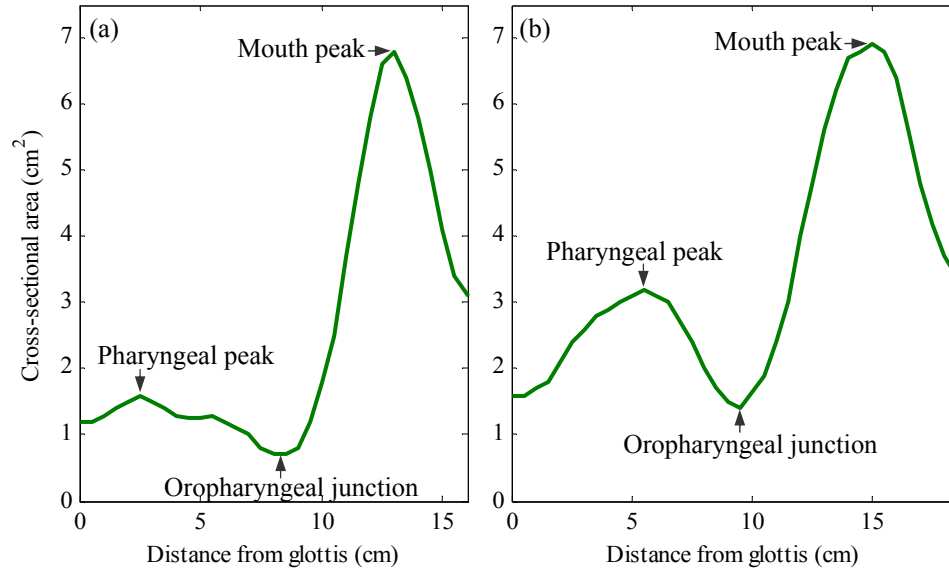
**Figure 8.6**   Upper airway (UA) area-distance profiles of reference (a) UA1 and (b) UA2.

calculated from the YIN algorithm [267], which is a meliorated autocorrelation-based pitch estimator with minimum estimation errors, by means of the cumulative mean normalized difference function

$$d_t^{'}(\tau) = \begin{cases} 1 & , \quad \text{if } \tau = 0 \\ d_t(\tau) \Big/ \left[ (1/\tau) \sum_{k=1}^{\tau} d_t(k) \right] & , \quad \text{otherwise} \end{cases} \tag{8.12}$$

where

$$d_t(\tau) = \sum_{k=t-\tau/2-N/2}^{t-\tau/2+N/2} (x_k - x_{k+\tau})^2 \tag{8.13}$$

is the difference function of lag $\tau$ at time index $t$, and $N$ is the frame size of a discrete-time snore signal $x_t$. The SFs were selected as the excitations for three different types of synthetic snores, one for each snore type.

Besides the selection of SF, we chose two clinical UA area-distance profiles from the works of Jung *et al.* [268] and Mohsenin [269] for acoustic modeling, as depicted in Figure 8.6 with familiar anatomical landmarks (e.g., pharyngeal peak, oropharyngeal junction, and mouth peak) labeled between the glottis and the incisors.
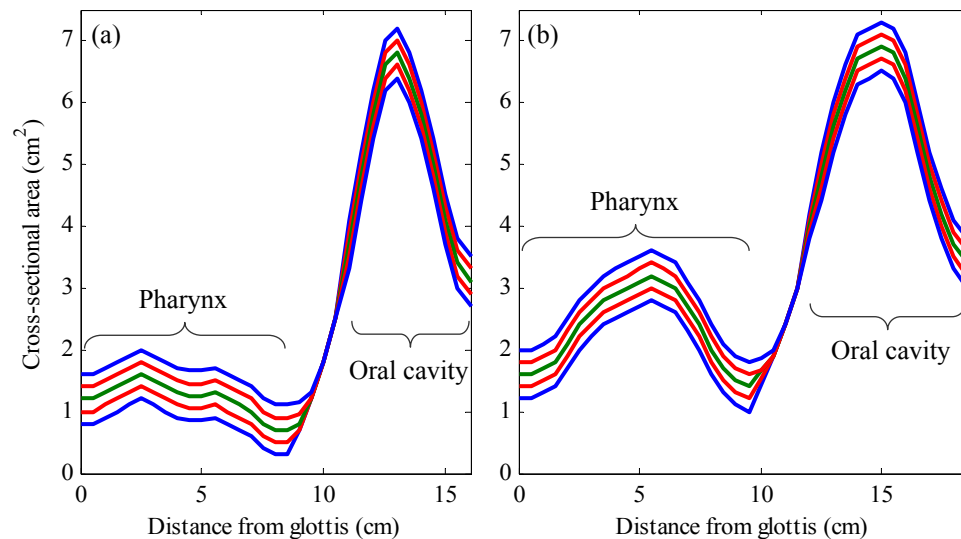
**Figure 8.7**    Upper airway (UA) area-distance profiles with changes of cross-sectional areas of pharynx and oral cavity at ±0.2 cm$^2$ (red) and ±0.4 cm$^2$ (blue) from the reference (a) UA1 (green) and (b) UA2 (green).

The profiles were constructed by the acoustic reflection technology and intended as reference UA1 and UA2, respectively; the former has a shorter UA (from glottis to incisors) length than the latter (around 16.0 cm versus 18.5 cm), possibly due to different subjects and measurement procedures. Alterations in the CSA of PX (from glottis to soft palate) and OC (from anterior margin of oropharyngeal junction to incisors) were made in ±0.2 cm$^2$ and ±0.4 cm$^2$ from the reference models, as exemplified in Figure 8.7. For instance, PX +0.4 cm$^2$ and OC -0.4 cm$^2$ specify a CSA increment of PX by 0.4 cm$^2$ and a decrement of OC by 0.4 cm$^2$ with respect to the reference, which in turn illustrates a subject with a smaller pharyngeal airway but wider mouth opening during snoring.

Using the above resources, snores were synthesized by convolving the SF and the UA transfer function with lip radiation load. For each UA model, 12 CSA perturbations were performed, thereby giving 39 synthetic snores including reference-generated snores of roughly 1.5 s each for analysis.

## 8.3.1.2    Test Procedure

Synthetic snores of the area perturbations were analyzed by means of the DAP modeling [173] and the Welch's averaged modified periodogram method [112,113] with the same parameters selected for the parametric analysis of natural snores in Section 5.5.2 (a 256-sample Hanning window with 75% overlap). F1 and PF were extracted via peak picking the 14th-order DAP spectrum and the Welch PSD spectrum, respectively.

On top of the acoustical analysis, six pairs of synthetic snores from each UA model were subjectively experimented by the group of 16 normal-hearing polysomnographic technicians and signal processing specialists through a paired comparison approach [232,245], one at a time, in a quiet room. For each pair of snore sounds, the listener was requested to compare one sound (comparison sample) with the other (reference sample), and then rate the comparison sample, in terms of loudness, sharpness, roughness, fluctuation strength, and annoyance, on a 5-point semantic differential scale with bipolar adjective pairs [232,245] before proceeding to the next pair.

For example, in the case of loudness, the value '1' in the scale symbolises very soft, '2': somewhat soft, '3': equal loudness, '4': somewhat loud, and '5': very loud. If the listener perceived the comparison sample as somewhat soft, a value of '2' will be assigned to it, and automatically a value of '4' will be assigned to the reference sample, implying that it is somewhat loud. Thus, the degree of loudness for the comparison sample and the reference sample are '2' and '4', correspondingly. In contrast, if the listener judged both the samples as equal loudness, a value of '3' will be allocated to the samples, and the degree of loudness for the samples is '3'.
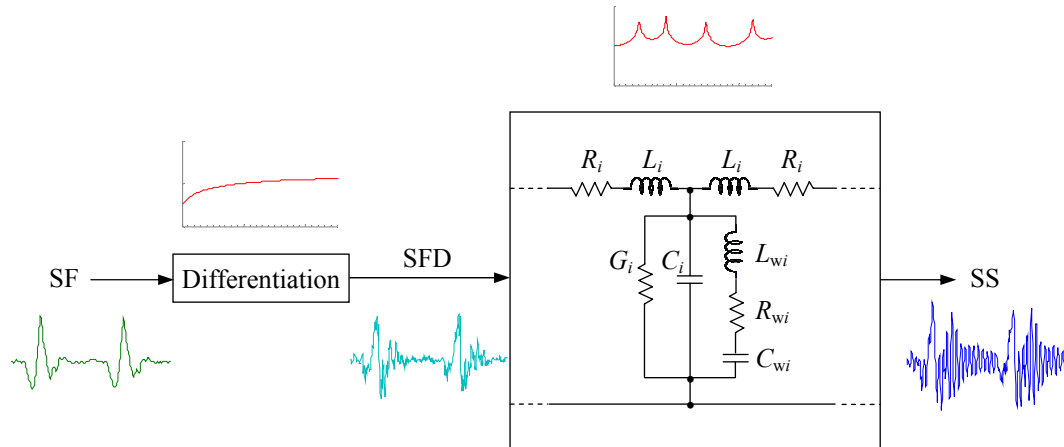
**Figure 8.8**   Source-filter model of snore production with snore source flow derivative (SFD) as the input to the upper airway transmission-line model with lumped parameters including resistance $R_i$, inertance $L_i$, compliance $C_i$, conductance $G_i$, wall inertance $L_{wi}$, wall resistance $R_{wi}$, and wall compliance $C_{wi}$. SF and SS denote source flow and synthetic snore, respectively.

### 8.3.2     Experimental Conditions for Source Flow Analysis

### 8.3.2.1     Signal Parameterization

Apart from the area perturbation study, the analysis of SF and SFD is also necessary for the understanding of snore production mechanisms, which can also be rationalized by the source-filter theory [152,158-161]. Aerodynamic forces at constrictions in the UA excite ES vibrations, modulating the steady airflow to a pulsating airflow (snore SF). The SF, denoted by $u(t)$, is then modified by the UA acting as an acoustic filter with impulse response of $h(t)$. The volume velocity output of the UA is eventually radiated from the lips, which can be regarded as a differentiator [152,158-161], giving rise to a snore sound pressure signal $x(t) \approx d[u(t) * h(t)]/dt = [du(t)/dt] * h(t)$, where $du(t)/dt$ is the SFD that combines the effects of SF and lip radiation and constitutes the input to the UA, as graphically displayed in Figure 8.8.

Figure 8.9 renders the SFs and SFDs of the abovementioned quasi-periodic snore signals. To establish quantitative data on the dynamical behavior of ES, we parameterized the SF and SFD waveforms in time domain after smoothing the
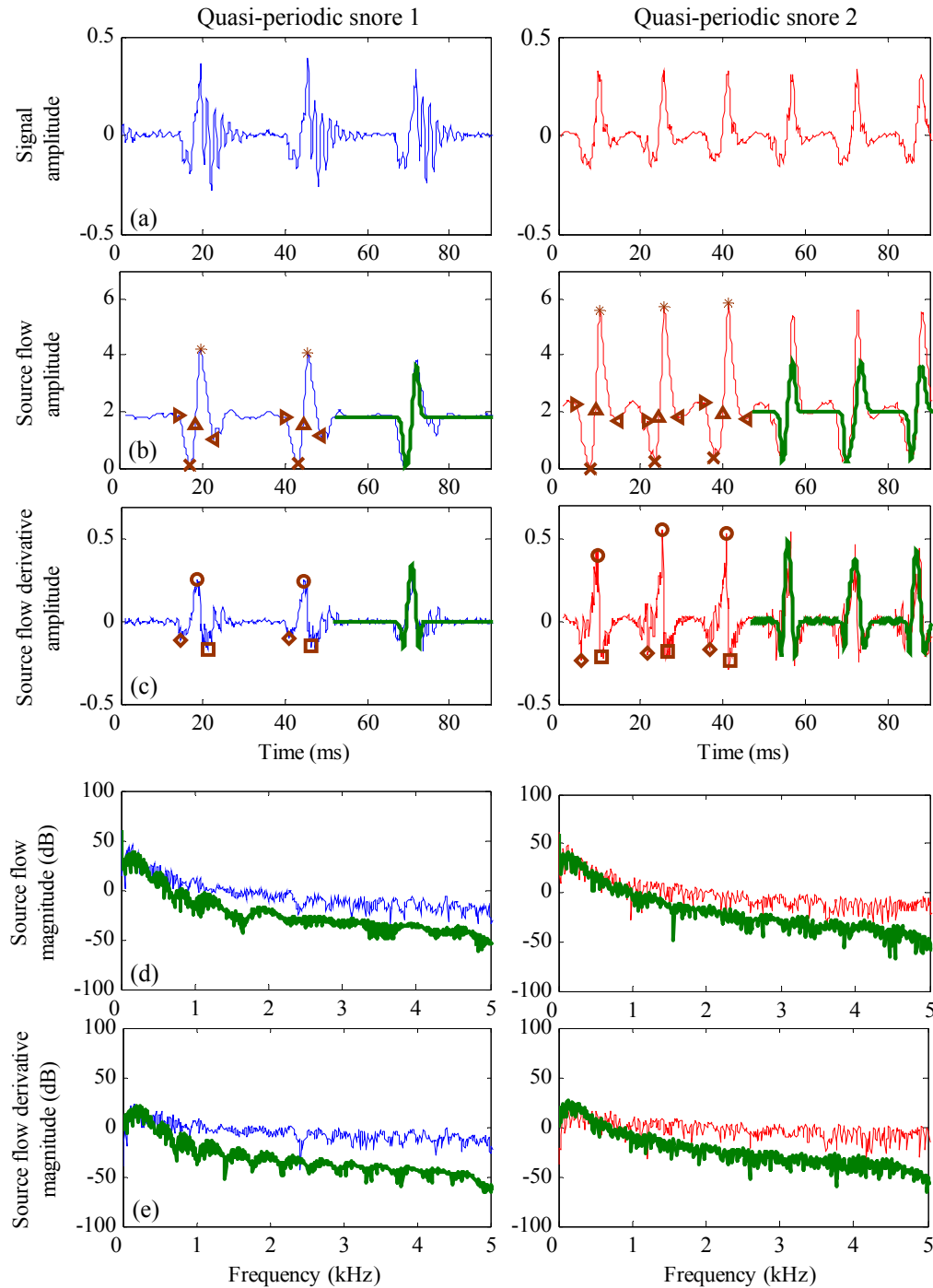
**Figure 8.9**  Waveforms of (a) quasi-periodic snores, and their corresponding (b) source flow (SF) waveforms, (c) source flow derivative (SFD) waveforms, (d) SF spectra, and (e) SFD spectra. Waveforms and spectra for SFD model with support of [-5, 5] are plotted in green, whereas time instants are indicated by markers.

waveforms to suppress short-lived fluctuations driven by interharmonic noise above 2000 Hz [160,270]. Figure 8.10 plots a common SF waveform and SFD waveform
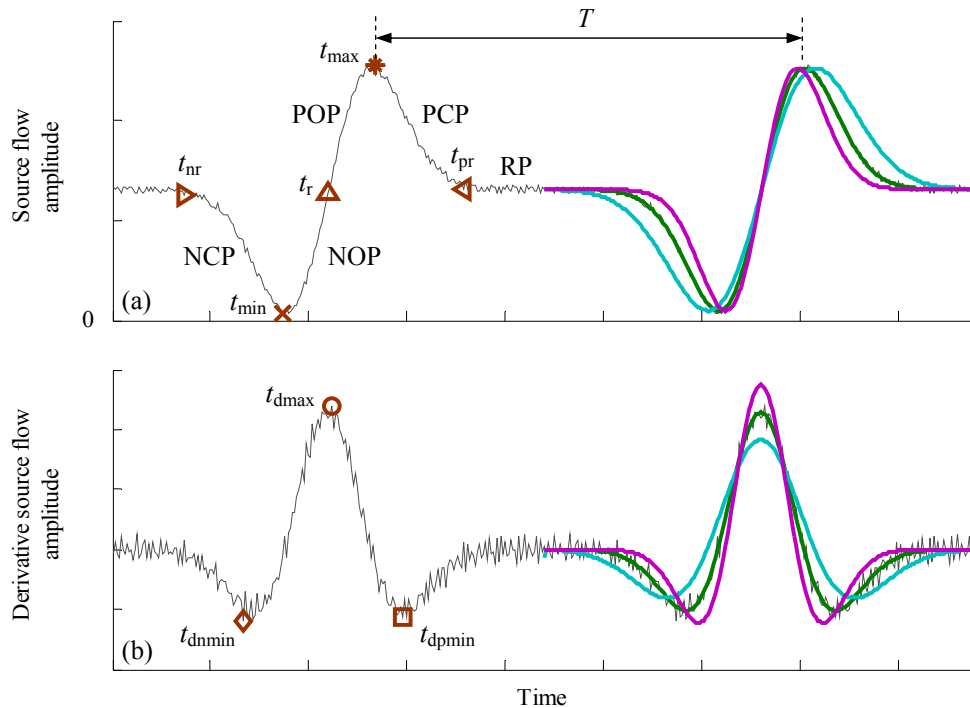
**Figure 8.10**   Common waveforms of snore (a) source flow and (b) source flow derivative with temporal features of function support intervals [-4, 4] (cyan), [-5, 5] (green), and [-6, 6] (magenta). $T$ denotes fundamental period; $t_{nr}$ and $t_{pr}$, negative and positive return flow time instant, respectively; $t_{min}$ and $t_{max}$, minimum and maximum flow time instant, respectively; $t_r$, return flow time instant; $t_{dnmin}$ and $t_{dpmin}$, negative and positive minimum flow derivative time instant, respectively; $t_{dmax}$, maximum flow derivative time instant; NCP and NOP, negative closing and negative opening phase, respectively; POP and PCP, positive opening and positive closing phase, respectively; RP, return phase.

with distinguishing time instants, such as negative and positive return flow time instants, minimum and maximum flow time instants, as well as return flow time instant, to define the onset of five plausible vibratory phases within a fundamental period. In sequence, these phases are negative closing and opening phases, positive opening and closing phases, and return phase. The time instants and their corresponding amplitude instants primarily serve as a basis for formulating several critical time- and amplitude-based parameters to quantify ES oscillatory maneuvers, as detailed in Table 8.2.

**Table 8.2**   Time- and amplitude-based parameters for snore source flow analysis.

| Parameter | Expression | Parameter | Expression |
|---|---|---|---|
| Open quotient | $OQ = \dfrac{t_{pr} - t_{nr}}{T}$ | Positive opening quotient | $POQ = \dfrac{t_{max} - t_{r}}{t_{pr} - t_{nr}}$ |
| Negative open quotient | $NQ = \dfrac{t_{r} - t_{nr}}{T}$ | Positive closing quotient | $PCQ = \dfrac{t_{pr} - t_{max}}{t_{pr} - t_{nr}}$ |
| Positive open quotient | $PQ = \dfrac{t_{pr} - t_{r}}{T}$ | Amplitude quotient | $AQ = \dfrac{1}{A_{dmax}}$ |
| Negative closing quotient | $NCQ = \dfrac{t_{min} - t_{nr}}{t_{pr} - t_{nr}}$ | Negative amplitude quotient | $NAQ = \dfrac{1}{\left| A_{dnmin} \right|}$ |
| Negative opening quotient | $NOQ = \dfrac{t_{r} - t_{min}}{t_{pr} - t_{nr}}$ | Positive amplitude quotient | $PAQ = \dfrac{1}{\left| A_{dpmin} \right|}$ |

$T$ refers to fundamental period; $t_{nr}$, negative return flow time instant (i.e., last negative zero-crossing of the flow derivative before time instant of negative minimum flow derivative); $t_{min}$, minimum flow time instant; $t_{r}$, return flow time instant; $t_{max}$, maximum flow time instant; $t_{pr}$, positive return flow time instant (i.e., first positive zero-crossing of the flow derivative after time instant of positive minimum flow derivative); $A_{dmax}$, amplitude at time instant of maximum flow derivative; $A_{dnmin}$, amplitude at time instant of negative minimum flow derivative; $A_{dpmin}$, amplitude at time instant of positive minimum flow derivative.

### 8.3.2.2   Signal Modeling

In addition to signal parameterization, we modeled the SF and SFD waveforms using analytical functions. By fitting the SF and SFD waveforms to various functional forms, we observed that the first and second derivatives of the Gaussian probability density function, expressed as

$$u(t) = t \exp\left( \frac{-t^2}{2} \right) \tag{8.14}$$

and

$$\frac{du(t)}{dt} = \left( 1 - t^2 \right) \exp\left( \frac{-t^2}{2} \right) \tag{8.15}$$

can appropriately represent the SF and SFD pulse shapes, respectively. Coincidentally, the function for SFD model is proportional to the Mexican hat wavelet [136,221] that has been demonstrated to effectively identify QRS complexes in electrocardiograms [271], calcium sparks in muscle cells [272], and dynamic characteristics of micro-

electro-mechanical structures [273]. Stemmed from the admissibility condition of the wavelet [136,221], one can infer that the net gain of airflow during a fundamental period $T$ is zero since the area of the SFD model is zero, i.e.,

$$\int_0^T \frac{du(t)}{dt} = 0 .$$ 

(8.16)

This deduction for maintaining an overall constant volume velocity of airflow at ES is in agreement with that made in the Liljencrants-Fant glottal model [158,159,274], a popular glottal flow derivative waveform simulating voice source for speech analysis and synthesis.

A primary consideration for the SFD model is its function support interval fitting the SFD pulsating pattern. Even though the Mexican hat wavelet possesses an effective compact support of [-5, 5] [136,221], we further assessed the ability of two other possible support intervals, namely [-4, 4] and [-6, 6], in curve-fitting the shape of SFD pulses by performing listening tests on the resulting synthetic snores. By doing so, we also examined the perceptual impact of changing the SFD pulse shapes on the generation of snores. Temporal features of the support intervals are displayed in Figure 8.10, indicating discrepancies between the intervals. To augment the naturalness of synthetic snores, the devised time domain parameters in Table 8.2 were applied to construct artificial SFD waveforms.

### 8.3.2.3    Test Procedure

The performance of SFD model was subjectively appraised via listening tests of the synthetic snores, which were created by convolving the SFD and the UA transfer function without lip radiation load. The SFDs involve SFD1 and SFD2 with various support intervals, whereas the UA models involve reference UA1 and UA2 without area perturbations. Pairs of snore sounds were presented, one at a time, to the same

**Table 8.3**   First formant frequency (F1) and spectral peak frequency (PF) computed with a frame size = 256 samples or 23 ms of snore signals synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC) for upper airway models (UA1 and UA2).

| $\Delta$CSA (cm$^2$) | | UA1 | | | | | | UA2 | | | | | |
| | | SF1 | | SF2 | | SF3 | | SF1 | | SF2 | | SF3 | |
| PX | OC | F1 | PF | F1 | PF | F1 | PF | F1 | PF | F1 | PF | F1 | PF |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 547 | 602 | 694 | 691 | 688 | 685 | 564 | 559 | 593 | 599 | 585 | 581 |
| -0.4 | 0 | 581 | 563 | 705 | 702 | 704 | 705 | 602 | 576 | 609 | 633 | 602 | 601 |
| -0.2 | 0 | 560 | 600 | 704 | 699 | 703 | 698 | 569 | 561 | 599 | 613 | 592 | 591 |
| +0.2 | 0 | 545 | 599 | 676 | 672 | 679 | 673 | 560 | 554 | 590 | 647 | 572 | 569 |
| +0.4 | 0 | 527 | 578 | 666 | 664 | 667 | 660 | 552 | 546 | 583 | 662 | 566 | 560 |
| 0 | -0.4 | 551 | 632 | 685 | 679 | 682 | 675 | 559 | 554 | 589 | 617 | 575 | 572 |
| 0 | -0.2 | 558 | 629 | 690 | 688 | 686 | 673 | 561 | 556 | 593 | 615 | 582 | 576 |
| 0 | +0.2 | 536 | 587 | 696 | 694 | 694 | 690 | 564 | 559 | 598 | 602 | 589 | 585 |
| 0 | +0.4 | 551 | 584 | 700 | 695 | 700 | 694 | 566 | 560 | 598 | 603 | 591 | 588 |
| -0.4 | +0.4 | 587 | 560 | 709 | 705 | 706 | 707 | 626 | 608 | 619 | 633 | 611 | 650 |
| -0.2 | +0.2 | 549 | 598 | 706 | 701 | 703 | 699 | 568 | 559 | 608 | 622 | 598 | 596 |
| +0.2 | -0.2 | 522 | 596 | 672 | 668 | 674 | 673 | 556 | 552 | 579 | 673 | 566 | 709 |
| +0.4 | -0.4 | 520 | 563 | 659 | 657 | 659 | 652 | 548 | 542 | 578 | 612 | 556 | 616 |

Values are presented as mean. First row indicates F1 and PF in hertz for reference UA1 and UA2 at their respective CSA, while others indicate F1 and PF under various changes in CSA from the reference. $\Delta$CSA refers to change in CSA perturbations (cm$^2$).

group of polysomnographic technicians and signal processing specialists in a quiet room. For each pair of sounds, the listener was requested to rate the degree of similarity [232,245] of one sound (comparison sample) relative to the other (reference sample) on a numbered scale, which ranges from 1 (least similar) to 5 (most similar), before advancing to the next pair. Finally, the ratings were averaged to give a single evaluation score for each sound pair. Due to the fact that the reference samples are snores generated by either natural or smoothed SFD waveforms, the capability of SFD model in generating quality snores was also explored. To confirm the reliability of the listeners' judgments, we deliberately paired each sound with itself (i.e., comparison and reference samples were identical) and randomly incorporated them into the tests.
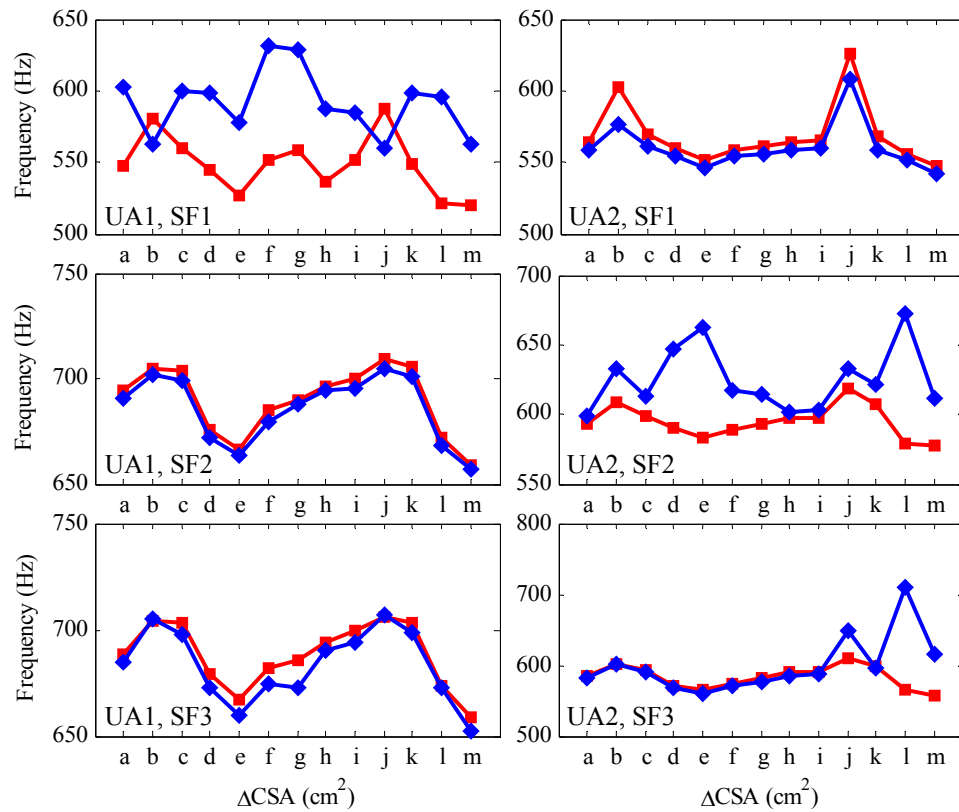
**Figure 8.11**   First formant frequency (red) and spectral peak frequency (blue) computed with a frame size = 256 samples or 23 ms of snore signals synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC) for upper airway models (UA1 and UA2). $\Delta$CSA denotes change in CSA perturbations ($cm^2$), while 'a' denotes (PX,OC) = (0,0); b = (-0.4,0); c = (-0.2,0); d = (+0.2,0); e = (+0.4,0); f = (0,-0.4); g = (0,-0.2); h = (0,+0.2); i = (0,+0.4); j = (-0.4,+0.4); k = (-0.2,+0.2); l = (+0.2,-0.2); m = (+0.4,-0.4).

Besides the subjective paired comparison scores, the sum-of-squared errors between the natural and stimulated SFD waveforms were computed to determine how well the latter fits the former. Apparently, the smaller the error, the better the stimulated waveform fits its natural counterpart.

## 8.4    Results and Discussion

### 8.4.1    Acoustical Influences of Area Perturbations

Table 8.3 and Figure 8.11 summarize the values of F1 and PF for perturbations that lie

**Table 8.4**  First formant frequency (F1) and spectral peak frequency (PF) computed with a frame size = 512 samples or 46 ms of snore signals synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC) for upper airway models (UA1 and UA2).

| $\Delta$CSA (cm$^2$) | | UA1 | | | | | | UA2 | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | SF1 | | SF2 | | SF3 | | SF1 | | SF2 | | SF3 | |
| PX | OC | F1 | PF | F1 | PF | F1 | PF | F1 | PF | F1 | PF | F1 | PF |
| 0 | 0 | 557 | 572 | 691 | 687 | 686 | 680 | 565 | 560 | 600 | 639 | 586 | 624 |
| -0.4 | 0 | 595 | 572 | 703 | 694 | 703 | 698 | 601 | 582 | 609 | 654 | 604 | 786 |
| -0.2 | 0 | 564 | 589 | 700 | 691 | 698 | 693 | 576 | 569 | 605 | 637 | 593 | 675 |
| +0.2 | 0 | 556 | 580 | 677 | 673 | 679 | 669 | 556 | 557 | 595 | 678 | 578 | 668 |
| +0.4 | 0 | 538 | 611 | 663 | 660 | 667 | 654 | 550 | 553 | 585 | 676 | 565 | 553 |
| 0 | -0.4 | 557 | 575 | 683 | 679 | 681 | 676 | 554 | 558 | 601 | 644 | 580 | 607 |
| 0 | -0.2 | 560 | 571 | 689 | 685 | 683 | 679 | 562 | 559 | 601 | 650 | 587 | 610 |
| 0 | +0.2 | 553 | 583 | 694 | 687 | 690 | 687 | 566 | 562 | 602 | 644 | 589 | 624 |
| 0 | +0.4 | 562 | 595 | 698 | 692 | 694 | 692 | 568 | 566 | 603 | 623 | 596 | 621 |
| -0.4 | +0.4 | 597 | 593 | 708 | 701 | 708 | 701 | 628 | 612 | 620 | 650 | 609 | 790 |
| -0.2 | +0.2 | 569 | 582 | 702 | 691 | 702 | 697 | 578 | 572 | 608 | 638 | 601 | 786 |
| +0.2 | -0.2 | 548 | 588 | 671 | 668 | 677 | 665 | 552 | 558 | 583 | 717 | 566 | 556 |
| +0.4 | -0.4 | 534 | 576 | 659 | 655 | 660 | 647 | 550 | 545 | 577 | 652 | 560 | 546 |

Values are presented as mean. First row indicates F1 and PF in hertz for reference UA1 and UA2 at their respective CSA, while others indicate F1 and PF under various changes in CSA from the reference. $\Delta$CSA refers to change in CSA perturbations (cm$^2$).

within ±0.4 cm$^2$ from the reference models. The key influence of decreasing the CSA of PX, while keeping constant or widening the CSA of OC, is an increase of F1. Correspondingly, an increase in the CSA of PX, while maintaining constant or narrowing the CSA of OC, can reduce F1. The variations of F1 with changing CSA are not as drastic as those of PF; conversely, they are more consistent and predictable for rise or fall regardless of SF types and UA models. In UA1 for SF2, when PX alters from -0.4 to +0.4 cm$^2$ in parallel with OC from +0.4 to -0.4 cm$^2$, both F1 and PF declines progressively from 709 to 659 Hz and from 705 to 657 Hz, respectively. However, under the same CSA configuration for SF2 in UA2, F1 drops from 619 to 578 Hz, but PF fluctuates between 612 and 673 Hz rather than decreases as before.
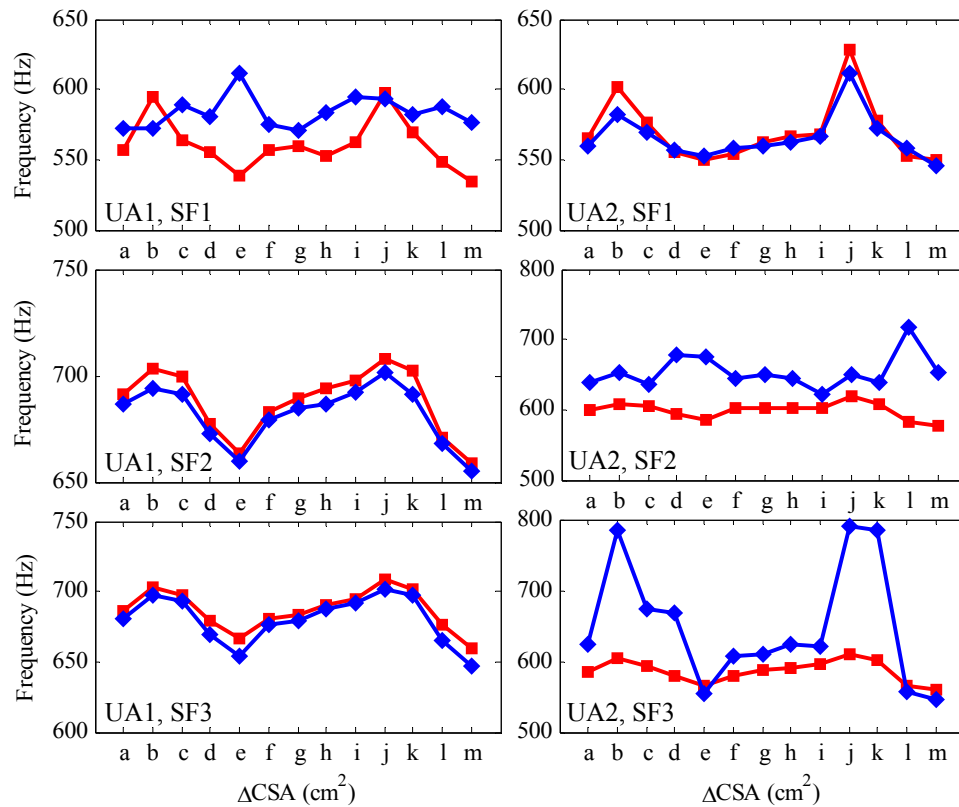
**Figure 8.12**    First formant frequency (red) and spectral peak frequency (blue) computed with a frame size = 512 samples or 46 ms of snore signals synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC) for upper airway models (UA1 and UA2). $\Delta$CSA denotes change in CSA perturbations (cm$^2$), while 'a' denotes (PX,OC) = (0,0); b = (-0.4,0); c = (-0.2,0); d = (+0.2,0); e = (+0.4,0); f = (0,-0.4); g = (0,-0.2); h = (0,+0.2); i = (0,+0.4); j = (-0.4,+0.4); k = (-0.2,+0.2); l = (+0.2,-0.2); m = (+0.4,-0.4).

Likewise, under the same CSA configuration in UA1 for SF3, both F1 and PF gradually decline from 706 to 659 Hz and from 707 to 652 Hz, respectively; however, in UA2 for SF3, while F1 drops from 611 to 556 Hz, PF fluctuates between 596 and 709 Hz. This undeterminable behavior of PF is also noticed in snore signals synthesized using SF1, and even when the frame size has been adjusted to 512 samples, as justified in Table 8.4 and Figure 8.12. A close inspection of Table 8.3 and Table 8.4 also highlight that increasing the CSA of OC or shortening the UA length can likely boost the values of F1.

The frequency marker F1 of snore signals has considerably proven to be more sensitive to changes in UA geometry than PF, which is in line with the previous speech and voice studies, recognizing the associations between F1 and constrictions in PX and OC: the greater the pharyngeal constriction or the lower the oral constriction, the higher is F1 [152,160,275]. For instance, the PX is more constricted and the lips are more broadly opened for the vowel /a/ than /i/; therefore, the value of F1 for /a/ (710 Hz) is higher than that of /i/ (280 Hz) [275]. With the establishment of anatomical-acoustical relationships, one can deduce that F1 is superior to PF in detecting OSA, further validating the reliability of F1, discussed in Chapter 5, as a potential diagnostic marker for OSA. Apneic patients tend to produce snores of higher F1 than that of benign patients because they usually possess smaller pharyngeal airway [29-31,56,57] and spend more time on oral or oro-nasal breathing than benign ones [276,277], further reducing the airway size as a consequence of jaw opening.

## 8.4.2    Perceptual Influences of Area Perturbations

The degree (range 1–5) of each psychoacoustic metric (loudness, sharpness, roughness, fluctuation strength, and annoyance) of snore sounds involved in the listening tests for the area perturbation study are tabulated in Table 8.5. Most listeners feel that the sound quality metrics in response to the CSA perturbations are not pronounced. The degree ratings for each pair of snore sounds with same SF, when compared across different SFs, are confounded, except for those tailored to investigate the effects of varying CSA of OC, which indicate no sound change. On the contrary, when we compared snore sounds of different SFs with the same CSA configuration, the degree ratings are conclusive. Sounds created by SF3, an aperiodic waveform, yield highest ratings for all the metrics (degree range 3.9–4.6), followed by the two quasi-periodic

**Table 8.5**    Psychoacoustic metrics, in terms of loudness (L), sharpness (S), roughness (R), fluctuation strength (F), and annoyance (PA), of snore sounds synthesized using different source flows (SF1, SF2, and SF3) and cross-sectional areas (CSA) of pharynx (PX) and oral cavity (OC).

| $\Delta$CSA (cm$^2$) | | SF1 | | | | | SF2 | | | | | SF3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PX | OC | L | S | R | F | PA | L | S | R | F | PA | L | S | R | F | PA |
| 0 | 0 | 1.4 | 1.5 | 1.7 | 1.7 | 1.5 | 3.2 | 2.8 | 3.1 | 3.4 | 2.9 | 4.4 | 4.6 | 4.2 | 3.9 | 4.6 |
| -0.4 | 0 | 2.8 | 2.9 | 3.1 | 3.1 | 3.1 | 3.0 | 3.3 | 3.3 | 2.8 | 3.0 | 3.3 | 3.1 | 3.4 | 3.1 | 3.3 |
| +0.4 | 0 | 3.2 | 3.1 | 2.9 | 2.9 | 2.9 | 3.0 | 2.7 | 2.7 | 3.2 | 3.0 | 2.7 | 2.9 | 2.6 | 2.9 | 2.7 |
| 0 | -0.4 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| 0 | +0.4 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| -0.4 | +0.4 | 2.7 | 2.8 | 3.0 | 2.8 | 2.9 | 3.0 | 2.7 | 3.0 | 3.0 | 2.8 | 3.3 | 3.0 | 3.3 | 3.3 | 3.2 |
| +0.4 | -0.4 | 3.3 | 3.2 | 3.0 | 3.2 | 3.1 | 3.0 | 3.3 | 3.0 | 3.0 | 3.2 | 2.7 | 3.0 | 2.7 | 2.7 | 2.8 |

Values are presented as mean for both upper airway models (UA1 and UA2) combined. First row indicates the degree (range 1 to 5, with '1' being least and '5' being most) of the metrics after comparing snore sounds with different SFs, while others indicate the degree of the metrics for each pair of snore sounds (row 2nd and 3rd, 4th and 5th, 6th and 7th) with same SF. $\Delta$CSA refers to change in CSA perturbations (cm$^2$).

waveforms, SF2 (degree range 2.8–3.4) and SF1 (degree range 1.4–1.7)

Unlike the frequency markers (F1 and PF), the integrity of snore sound qualities (loudness, sharpness, roughness, fluctuation strength, and psychoacoustic annoyance), are not easily affected by the changes of cross-sectional airway dimensions. Apart from the limited availability of listeners, another probable explanation is that the spectral contents (e.g., envelope, magnitude, and formant bandwidth) of the acoustic transfer functions of UA models with or without CSA perturbations have no considerable discrepancy. As evident in Figure 8.13, the transfer functions of reference UA models and those after CSA perturbations are somewhat similar, except for their spectral envelopes that are slightly shifted sideways. On the other hand, a change in SF types can bring about a vast change in the perception of snore sounds, suggesting that the UA anatomical structures are probably the aggravating factors in snore production.

The dynamics of SF (e.g., amplitude envelope, fundamental frequency, and harmonicity) contribute a dominant part to the psychoacoustics of snore sounds since
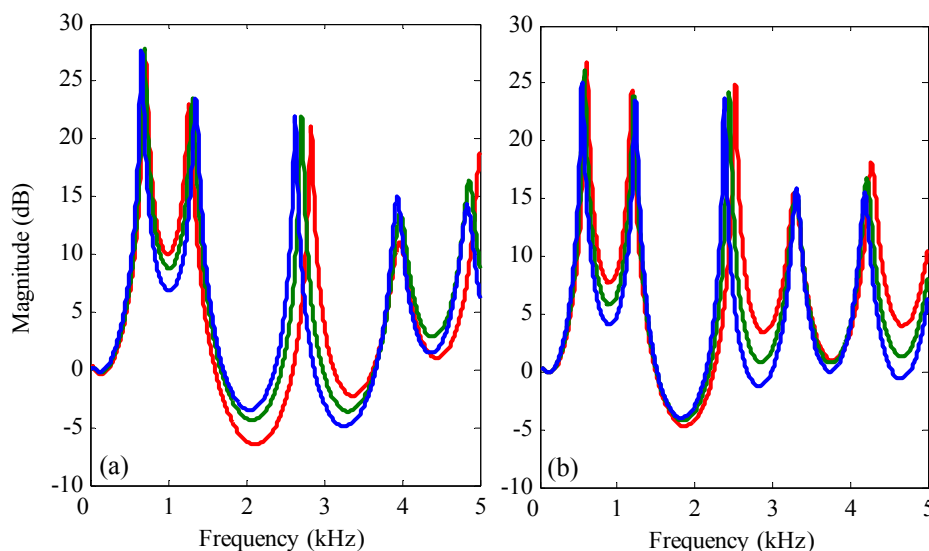
**Figure 8.13**   Acoustic transfer functions of reference (a) UA1 (green) and (b) UA2 (green), and transfer functions of models with cross-sectional areas (CSA) of pharynx (PX) at -0.4 cm$^2$ and oral cavity (OC) at +0.4 cm$^2$ from the reference (red), and models with CSA of PX at +0.4 cm$^2$ and OC at -0.4 cm$^2$ from the reference (blue).

each sound quality metric can be anticipated from acoustic measurements [232], as elaborated in Section 7.1. Loudness correlates with amplitude; sharpness depends on centre frequency and bandwidth; roughness and fluctuation strength govern by temporal variations of high-frequency and low-frequency, respectively; and annoyance relies on the above four metrics. Relating this to the mechanisms of snoring development [28,150,213-216], a narrower airway escalates the negative intraluminal pressure and hence enhances the driving forces on the vibrating soft tissues or the rate of pharyngeal airway closure and reopening. Spontaneously, the complexity of SF dynamics rises (e.g., amplifying the amplitude, increasing the fundamental frequency, and introducing higher harmonics), and the resulting sound waves are spectrally modified by the acoustic transfer function of UA, producing snoring sounds of build-up loudness, sharpness, roughness, fluctuation strength, and/or psychoacoustic annoyance. The perceptual findings further elucidate why AS and BS can be classified

via the psychoacoustics of snore sounds in Chapter 7.

### 8.4.3　　Source Flow Parameterization

Table 8.6 reveals the time- and amplitude-based parameter values for SF1 and SF2. Owing to the lengthened airflow plateau of SF1, the parameter values concerning the relative timing of ES maneuver within the fundamental period (open quotient, OQ; negative open quotient, NQ; and positive open quotient, PQ) are significantly smaller for SF1 than for SF2. The values of OQ $= 0.3203 \pm 0.0072$ (range 0.3139–0.3280) for SF1, whereas OQ $= 0.5706 \pm 0.0733$ (range 0.4743–0.6413) for SF2. Comparisons between NQ and PQ for each SF show that their values fall within the same range for SF1 but not for SF2 where NQ ($0.2465 \pm 0.0438$) are lesser than PQ ($0.3241 \pm 0.0380$). This outcome for SF2 is justified by the values of negative closing quotient (NCQ), which are considerably reduced as opposed to that of positive closing quotient (PCQ), with NCQ $= 0.2632 \pm 0.0485$ and PCQ $= 0.4250 \pm 0.0426$. Thus, the outcome suggests that the ES for SF2 takes a shorter time to collapse from neutral position than from fully open to neutral position, as evidenced in Figure 8.9. In contrast, such speed imbalance at the negative and positive closing phases is not found in the ES for SF1, signifying dissimilar biomechanical properties (e.g., compliance and elasticity) between the ESs.

In view of the amplitude findings, the values of amplitude quotient (AQ), interpreted as the inverse of maximum SFD amplitude, for SF2 (AQ $= 2.3278 \pm 0.2799$, range 1.9802–2.6316) are substantially lower than for SF1 (AQ $= 3.0173 \pm 0.1566$, range 2.8548–3.1672), which can be attributed to the rapid acceleration of SF2 as the ES abruptly pops open from closing position. Moreover, the values of positive amplitude quotient (PAQ) are consistently smaller than that of negative amplitude

**Table 8.6**  Time- and amplitude-based parameters for snore source flows, SF1 ($T = 26.2$ ms) and SF2 ($T = 15.6$ ms).

| Parameter | SF1 | | SF2 | |
|---|---|---|---|---|
| | Mean ± SD | Range | Mean ± SD | Range |
| $OQ = \dfrac{t_{pr} - t_{nr}}{T}$ | 0.3203 ± 0.0072 | 0.3139–0.3280 | 0.5706 ± 0.0733 | 0.4743–0.6413 |
| $NQ = \dfrac{t_{r} - t_{nr}}{T}$ | 0.1565 ± 0.0158 | 0.1448–0.1744 | 0.2465 ± 0.0438 | 0.1899–0.2936 |
| $PQ = \dfrac{t_{pr} - t_{r}}{T}$ | 0.1639 ± 0.0211 | 0.1396–0.1778 | 0.3241 ± 0.0380 | 0.2583–0.3552 |
| $NCQ = \dfrac{t_{min} - t_{nr}}{t_{pr} - t_{nr}}$ | 0.3409 ± 0.0342 | 0.3127–0.3789 | 0.2632 ± 0.0485 | 0.2037–0.3333 |
| $NOQ = \dfrac{t_{r} - t_{min}}{t_{pr} - t_{nr}}$ | 0.1482 ± 0.0256 | 0.1267–0.1766 | 0.1674 ± 0.0490 | 0.1279–0.2474 |
| $POQ = \dfrac{t_{max} - t_{r}}{t_{pr} - t_{nr}}$ | 0.1621 ± 0.0163 | 0.1435–0.1738 | 0.1444 ± 0.0214 | 0.1187–0.1839 |
| $PCQ = \dfrac{t_{pr} - t_{max}}{t_{pr} - t_{nr}}$ | 0.3489 ± 0.0414 | 0.3011–0.3731 | 0.4250 ± 0.0426 | 0.3608–0.4759 |
| $AQ = \dfrac{1}{A_{dmax}}$ | 3.0173 ± 0.1566 | 2.8548–3.1672 | 2.3278 ± 0.2799 | 1.9802–2.6316 |
| $NAQ = \dfrac{1}{\lvert A_{dnmin} \rvert}$ | 7.6117 ± 1.1294 | 6.4722–8.7306 | 5.5982 ± 0.9090 | 4.8190–7.2935 |
| $PAQ = \dfrac{1}{\lvert A_{dpmin} \rvert}$ | 4.3133 ± 0.4390 | 3.8202–4.6615 | 4.6915 ± 0.3747 | 4.2766–5.2641 |

Values are presented as mean ± standard deviation (SD) and range. SF refers to source flow; $T$, fundamental period; OQ, open quotient; NQ, negative open quotient; PQ, positive open quotient; NCQ, negative closing quotient; NOQ, negative opening quotient; POQ, positive opening quotient; PCQ, positive closing quotient; AQ, amplitude quotient; NAQ, negative amplitude quotient; PAQ, positive amplitude quotient; $t_{nr}$, negative return flow time instant (i.e., last negative zero-crossing of the flow derivative before time instant of negative minimum flow derivative); $t_{min}$, minimum flow time instant; $t_{r}$, return flow time instant; $t_{max}$, maximum flow time instant; $t_{pr}$, positive return flow time instant (i.e., first positive zero-crossing of the flow derivative after time instant of positive minimum flow derivative); $A_{dmax}$, amplitude at time instant of maximum flow derivative; $A_{dnmin}$, amplitude at time instant of negative minimum flow derivative; $A_{dpmin}$, amplitude at time instant of positive minimum flow derivative.

quotient (NAQ) for both SF1 (PAQ = 4.3133 ± 0.4390; NAQ = 7.6117 ± 1.1294) and

SF2 (PAQ = 4.6915 ± 0.3747; NAQ = 5.5982 ± 0.9090), implying a relatively steep

decline in airflow when ES falls from fully open to neutral position at the positive closing phase. However, due to the large change of airflow, which is about two times greater, in the positive closing phase than in the negative closing phase, the time taken for ES to reach neural position is longer than it takes to collapse, as noted in the earlier outcome for SF2.

The mechanisms of snore production are often associated with the instability of airflow over the flexible soft tissues [28,150,213-216], and they can be explained by the theory of flutter and the concept of static divergence, in which the former is mainly responsible for palatal snoring (e.g., sound from soft palate vibration), while the latter is for pharyngeal snoring (e.g., sound from pharyngeal wall vibration) [215,216]. In a nutshell, when palatal snoring is triggered, the soft palate flaps backwards and forwards in the UA, briefly occluding the nasopharyngeal space for nasal breathers, or alternately obstructing the oral and nasal airways as it flutters between the tongue and the posterior pharyngeal wall for ora-nasal breathers. On the contrary, pharyngeal snoring involves airway wall oscillations around a neutral position, colliding with the opposing walls between the soft palate and the larynx.

Regardless of any modes of snoring, the UA lumen is inevitably susceptible to a series of short closures (partially or fully) and reopenings attributable to the flow-induced self-sustained vibration of soft tissues that are aeroelastically unstable during snoring [28,99,100,150,208,213-216,255-257]. According to the water hammer phenomenon [278], the airway internal pressure rises during each closure and then blasts open the shut lumen at an elevated speed, momentarily distending the airway and inducing an overshoot [215]. The overshoot effect may clarify why the amplitude of SF1 and SF2 in Figure 8.9 was intensified at the positive opening phase, followed by a decay indicating the return of airflow to its neutral value at the positive closing

phase. This closing-opening-return sequence characterizes the vibration rate of ES and equivalently determines the fundamental frequency of the quasi-periodic snores. The faster the vibration rate, the higher the fundamental frequency, and vice versa.

Relating this to the biomechanical properties of soft tissues, the ability for a tissue to stretch and return after stretching heavily relies on its compliance and elasticity, which vary among tissues and degree of tissue abnormalities [279-283]. Electromyographic studies have also reported that stiffness (inverse of compliance) of the soft palate for adult apneic subjects is higher than that of the tongue, whereas elasticity of the tongue is lower than that of the soft palate [284,285]. Furthermore, apneic subjects exhibit increased stiffness but decreased elasticity of the tongue as compared to non-snorers [286,287]. Surgical treatments for palatal snoring (e.g., injection snoreplasty [288,289], laser assisted uvulopalatoplasty [290,291], and palatal implants [292,293]) have also been documented to raise tissue stiffness, reducing dynamic flutter and snore loudness. Therefore, one may reasonably infer that the temporal details of SF waveform, quantified by the time- and amplitude-based parameters, carry rich information about the vibratory properties of snore ES and is potentially helpful for pathological research on snoring.

### 8.4.4    Source Flow Modeling

The paired comparison scores of synthetic snores generated from the SFD model with various function support intervals are tabulated in Table 8.7. As can be seen, the support of [-5, 5] and [-6, 6] yield comparable high ratings on the degree of similarity for all possible paired comparisons, while the support of [-4, 4] yields the lowest ratings. Since the artificial SFD waveforms were simulated using the post-smoothing time domain parameters, ratings for the comparisons between the synthetic and

**Table 8.7**    Paired comparison scores of snores synthesized at various function support intervals for different source flow derivatives (SFD1 and SFD2) and upper airway models (UA1 and UA2).

| Type of sound sample | | UA1 | |
| --- | --- | --- | --- |
| Comparison | Reference | SFD1 | SFD2 |
| NS | NS | 4.8125 ± 0.4031 | 4.7500 ± 0.4472 |
| $NS_m$ | NS | 4.0769 ± 0.2774 | 3.5385 ± 0.6602 |
| $SS_4$ | NS | 2.3846 ± 0.8697 | 1.5385 ± 0.5189 |
| $SS_5$ | NS | 2.7692 ± 0.5991 | 1.8462 ± 0.8987 |
| $SS_6$ | NS | 2.8462 ± 0.5547 | 1.8462 ± 0.8987 |
| $NS_m$ | $NS_m$ | 4.9375 ± 0.2500 | 4.8750 ± 0.3416 |
| $SS_4$ | $NS_m$ | 3.0000 ± 0.9129 | 2.6154 ± 0.9608 |
| $SS_5$ | $NS_m$ | 3.7692 ± 0.7250 | 3.0000 ± 0.8165 |
| $SS_6$ | $NS_m$ | 3.7692 ± 0.5991 | 3.0000 ± 0.8165 |
| $SS_4$ | $SS_4$ | 5.0000 ± 0.0000 | 5.0000 ± 0.0000 |
| $SS_5$ | $SS_5$ | 5.0000 ± 0.0000 | 5.0000 ± 0.0000 |
| $SS_6$ | $SS_6$ | 5.0000 ± 0.0000 | 5.0000 ± 0.0000 |

| Type of sound sample | | UA2 | |
| --- | --- | --- | --- |
| Comparison | Reference | SFD1 | SFD2 |
| NS | NS | 4.8750 ± 0.3416 | 4.8125 ± 0.4031 |
| $NS_m$ | NS | 4.0000 ± 0.4082 | 3.3846 ± 0.6504 |
| $SS_4$ | NS | 2.3077 ± 0.6304 | 1.9231 ± 0.6405 |
| $SS_5$ | NS | 2.8462 ± 0.6887 | 1.9231 ± 0.6405 |
| $SS_6$ | NS | 2.8462 ± 0.6887 | 1.9231 ± 0.4935 |
| $NS_m$ | $NS_m$ | 4.9375 ± 0.2500 | 4.8750 ± 0.3416 |
| $SS_4$ | $NS_m$ | 2.9231 ± 0.9541 | 2.2308 ± 0.7250 |
| $SS_5$ | $NS_m$ | 3.2308 ± 0.9268 | 2.6154 ± 0.7679 |
| $SS_6$ | $NS_m$ | 3.3077 ± 0.7511 | 2.6923 ± 0.7511 |
| $SS_4$ | $SS_4$ | 5.0000 ± 0.0000 | 5.0000 ± 0.0000 |
| $SS_5$ | $SS_5$ | 5.0000 ± 0.0000 | 5.0000 ± 0.0000 |
| $SS_6$ | $SS_6$ | 5.0000 ± 0.0000 | 5.0000 ± 0.0000 |

Values are presented as mean ± standard deviation. SFD refers to source flow derivative; UA, upper airway; NS, natural snore; $NS_m$, snore generated from smoothed source flow derivative; $SS_k$, synthetic snore with function support interval [-k, k].

smoothed snores are elevated. Snore sounds created from UA1 and SFD1 with support intervals [-4, 4], [-5, 5], and [-6, 6] are respectively rated with 2.3846 ± 0.8697, 2.7692 ± 0.5991, 2.8462 ± 0.5547 when they were evaluated against those generated from

natural SFD1. Conversely, the correspondingly ratings noticeably increase to $3.0000 \pm 0.9129$, $3.7692 \pm 0.7250$, and $3.7692 \pm 0.5991$ when the same snore sounds were weighted against those from smoothed SFD1. The listeners' judgments are reliable as the average paired comparison scores for self-pairing are persistently skewed towards the high end (at least 4.75 out of 5).

Although the SFD model with support of [-6, 6] somewhat outperforms the other support intervals in the listening tests, the sum-of-squared error for [-6, 6] is the largest, followed by [-5, 5] and [-4, 4], in a descending order (for SFD1: 3.5163, 3.3896, and 3.1408; and for SFD2: 16.0063, 15.9262, and 15.9062). The large error is plausibly a consequence of narrowness of temporal pulse, which is inversely proportional to the support interval width, as illustrated in Figure 8.10. By considering the trade-off between degree of similarity and sum-of-squared error, we recommended the support interval [-5, 5] for the SFD model, which is also the effective compact support for the Mexican hat wavelet [136,221]. Figure 8.9 depicts the curve-fitting of the SF and SFD waveforms using the SFD model with support of [-5, 5], along with their corresponding spectra.

Spectra of the natural, smoothed, and synthetic snore signals for different combinations of SFDs and UA models in Figure 8.14 offer extra confirmation of the earlier listening test results for SF analysis. The spectrum envelopes of synthetic snore signals match more closely to that of the smoothed signals than that of the natural signals. A close inspection of the plots also advocates that the proposed SFD model is capable of producing human-like snores up to 500 Hz and 1000 Hz for SFD1 waveform and SFD2 waveform, respectively. With these outcomes for various function support intervals modeling the SFD, one can realize that there exist acoustical and perceptual impacts of changing the shapes of SF or SFD pulse on snore quality.
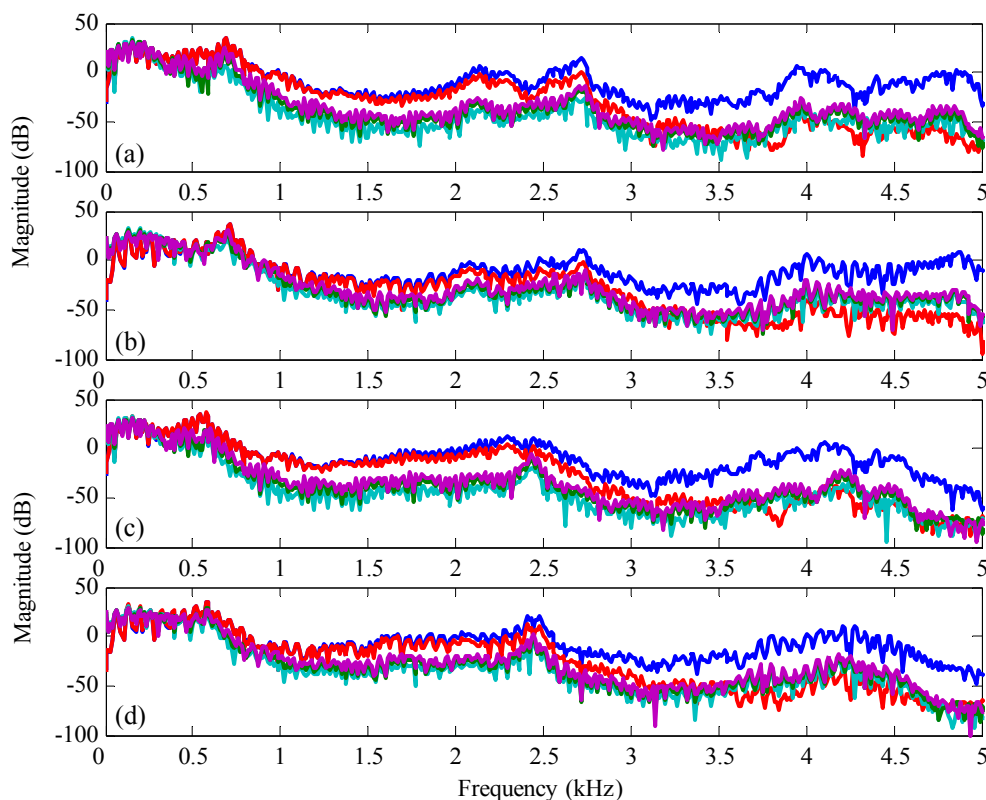
**Figure 8.14** Spectra of natural (blue), smoothed (red), and synthetic snores with function support intervals [-4, 4] (cyan), [-5, 5] (green), and [-6, 6] (magenta) for different combinations of source flow derivatives and upper airway models: (a) SFD1 and UA1; (b) SFD2 and UA1; (c) SFD1 and UA2; and (d) SFD2 and UA2.

Even though the biophysics of snore and speech production shares many similarities (e.g., source-filter model) [99,162], snore SF and glottal flow can be dissimilar because of their sound-generating physiology and mechanisms. As rendered in Figure 8.15, the Liljencrants-Fant glottal model [158,159,274] suitably fits the glottal flow derivative waveforms for vowels /a/ and /o/, which were acquired and processed in the same way as the snores, but not the snore SFD waveform. In contrast, the second derivative of the Gaussian function or the Mexican hat wavelet [136,221] is relatively superior in approximating the shape of SFD pulses, producing close-to-natural sounding snores.
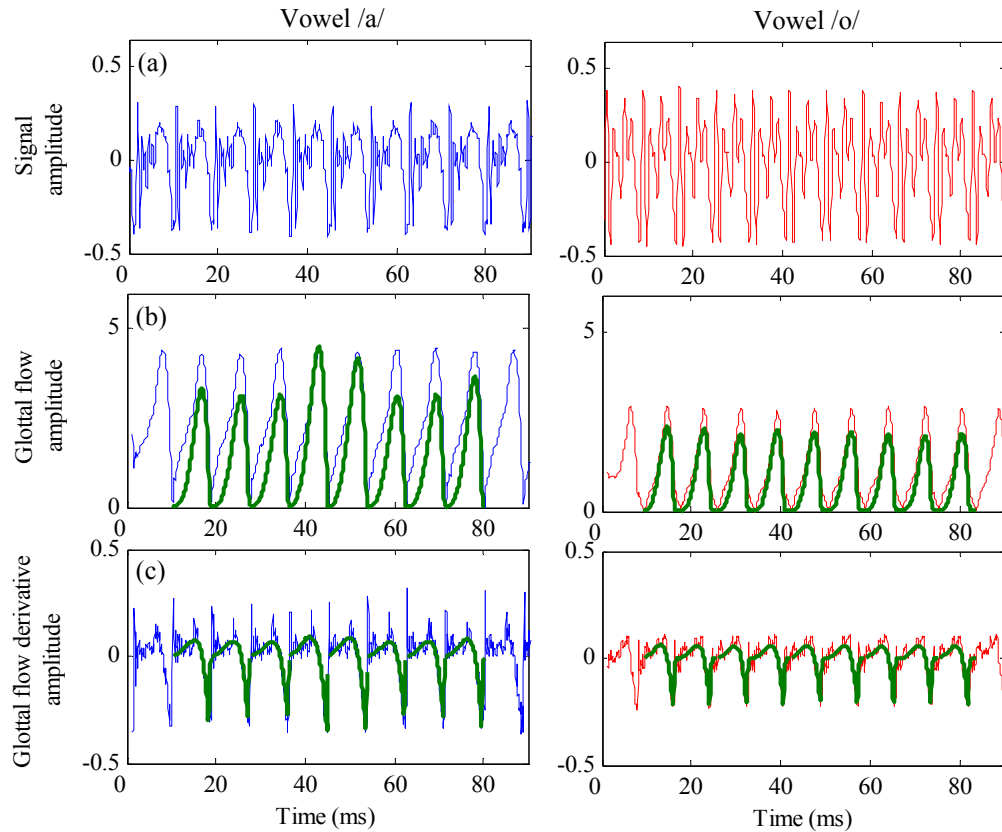
**Figure 8.15**     Waveforms of (a) vowels /a/ and /o/, and their corresponding (b) glottal flow waveforms and (c) glottal flow derivative waveforms. Waveforms for Liljencrants-Fant glottal model are plotted in green.

## 8.5    Summary

This chapter investigates the relationships between the snore SF, the UA anatomical structures, and the acoustical and perceptual characteristics of snores by conducting an area perturbation study and a SF analysis. Results demonstrate that changes in the CSA of PX and OC have diverse implications for generating snores, more acoustically than perceptually. The values of F1 and occasionally PF of snore signals increase with narrowing pharyngeal airway, being higher for a larger mouth opening, warranting the likelihood of using acoustical signatures in snores to discriminate between patients with and without OSA. Conversely, the snore sound psychoacoustic metrics are not easily affected by increasing or decreasing the dimensions of pharyngeal airway and

mouth opening; nonetheless, these CSA changes can indirectly influence the metrics by altering the waveforms of SF.

In addition to the above findings, changes in the SF or SFD pulse shape can affect the snore characteristics, both acoustically and perceptually. The shapes of SF pulse are dissimilar among snores and can be associated with the dynamic biomechanical properties (e.g., compliance and elasticity) of ES, lending support to our hypothesis that attributes of SF and SFD may contain important information about ES dynamics. The proposed SFD model based on the second derivative of the Gaussian probability density function or the Mexican hat wavelet with effective compact support of [-5, 5] can successfully generate natural-sounding synthetic snores, providing a deeper understanding of snore production mechanisms from physiological to acoustical and perceptual aspects.

Although this research has achieved its overall aim of implementing a convenient, low-cost, safe, and efficient approach for OSA screening through the acquisition and analysis of snore signals, one also has to acknowledge limitations of this research work. The next chapter of this thesis revisits the general aim and specific objectives of this work, concludes the research findings, and outlines the work limitations, along with recommendations for future research.

Chapter 9

# Conclusions and Recommendations

This research aims to innovate on a simple, inexpensive, safe, and reliable approach to diagnose OSA via the acquisition and analysis of snore signals, with an intention to mass screen for OSA. The present chapter revisits the objectives of this research work, summarizes the findings pertaining to the objectives, and draws conclusions from the findings. Subsequently, we discussed the work limitations and offered recommendations for future research.

## 9.1    Conclusions

OSA is a common sleep-related breathing disorder with elevated undiagnosis and untreated rate. Snore-based analysis has been increasing studied for its potential application in the clinical assessment of OSA as it can provide convenient and non-invasive data measurement with minimum medical and labor costs. Motivated by the advantages of processing snores, several snore-based diagnostic methodologies, in terms of snore counting, sound intensity measurement, power spectrum estimation, and pitch calculation, have been introduced to distinguish between patients with and without OSA (see Chapter 2). While corroborating the hypothesis that snore signals could be an alternative means for assessing OSA, these methodologies yield mixed, contradictory, and inconclusive outcomes, thereby highlighting the need for better approaches to diagnose OSA. We attempted to address this concern through accomplishing the following objectives.

**(a)    To develop and set up an acquisition system for snore signals**

A high-fidelity and user-friendly acquisition system for snore signals is designed and implemented in a clinical environment (see chapter 3). In the process of developing the acquisition system, we took into account of various design considerations for microphone, preamplifier, data acquisition card, grounding layout, graphical user interface, and MMD. Experiments were also conducted to warrant the acquisition system robustness; results show that snore signals exhibit strong immunity to background acoustical noise with correct selection of the amplification gain and improve rejections to EMI with careful grounding and shielding. A MMD of 0.3 m can also reach a good balance between signal quality and patient comfort. Therefore, an appropriate selection and installation of the system components is necessary to acquire snore signals of high quality and integrity.

**(b)    To improve snore signal quality and detect snore activity**

An advanced wavelet-driven preprocessing system is devised to simultaneously enhance snore signal quality via a LCD threshold and identify SA using a SA detector in a TIDWT domain (see Chapter 4). The LCD threshold with hard-thresholding rule successfully reduces additive colored noise embedded in snore signals, with an increased SNR of 3.02–38.22 dB and MOS of 4.14 out of 5.00, outperforming the classical LD threshold. The LCD threshold statistical quality was also verified theoretically. In addition, the SA detector is superior to the conventional short-time energy and zero-crossing rate approaches in identifying snore segments, with an accuracy of 50.63–97.47%. We also examined the clinical usefulness of the preprocessing system, and results reveal a better OSA diagnostic accuracy for snores analyzed after preprocessing than those before preprocessing. Thus, a robust system

for preprocessing snore signals is essential to improve signal quality prior to signal analysis and classification, augmenting the diagnostic value of snore signals for discriminating apneic patients from benign ones.

**(c)     To identify useful acoustical characteristics of snore signals as OSA diagnostic markers**

**(d)     To classify snores produced by apneic and benign patients using the diagnostic markers**

**(e)     To ascertain the correlation between AHI and the diagnostic markers**

A variety of acoustical signatures in snore signals, such as formant frequencies (see Chapter 5), WBC peaks (see Chapter 6), and psychoacoustic metrics (see Chapter 7), are proposed. F1, PF1, and PSF of snore signals can potentially be OSA diagnostic markers with high sensitivity (F1: 82.5–100%; PF1: 85.4–98.3%; PSF: 82.5–100%) and high specificity (F1: 82.0–95.0%; PF1: 85.0–100%; PSF: 83.3–100%) for all three patients groups (males, females, and both males and females combined), outperforming the commonly used spectral PF (sensitivity = 62.5–91.7%, specificity = 70.0–97.5%) in the classification of AS and BS (p-value < 0.0001). Besides that, loudness and annoyance of snore sounds can also discriminate between AS and BS, with a sensitivity of 72.0–78.0%, a specificity of 91.2–92.0%, and a p-value < 0.0001.

Accordingly, the possible relationship between AHI and the objective diagnostic markers (F1, PF1, and PSF) can best describe by the exponential or the power regression model with a predictive AHI value (F1: 10.3–14.9 events/h; PF1: 8.0–9.8 events/h; PSF: 9.3–10.7 events/h) close to the ideal value of 10.0 events/h. Furthermore, the correlation between AHI and the subjective diagnostic markers (loudness and annoyance) are found to be statistically significant (p-value < 0.0001)

with Pearson's product-moment correlation coefficient = 0.7182–0.7432 and Spearman's rank correlation coefficient = 0.7062–0.7162.

**(f)     To investigate physiological-anatomical-acoustical relationships of snores for validation of the diagnostic markers and for understanding the mechanisms of snore production**

The relationships between the snore SF, the UA anatomical structures, and the acoustical and perceptual characteristics of snores were explored by performing an area perturbation study and a SF analysis (see Chapter 8). Results illustrate that narrowing the pharyngeal airway consistently increases F1 but not PF; and altering the airway dimensions yield no considerable differences in perception of snore sounds but can ultimately affect the psychoacoustics by changing the dynamics of snore SF. Adding weight to these findings, alterations in the shape of SF or SFD pulse can influence the snore attributes, both acoustically and perceptually. The SF pulse shape varies among snores and can be related to the dynamic biomechanical properties (e.g., compliance and elasticity) of ES. The proposed SFD model stemmed from the second derivative of the Gaussian probability density function or the Mexican hat wavelet with effective compact support of [-5, 5] can competently generate close-to-natural sounding snores. Correlation exists between the snore SF, the UA anatomical structures, and the characteristics of snores. Hence, the proposed OSA diagnostic markers are reliable of classifying snores produced by apneic and benign patients, effectively distinguishing apneic patients from benign ones.

## 9.2    Recommendations for Further Research

Even though this research has accomplished its objectives emphasizing on the acquisition and analysis of snore signals for the diagnosis of OSA, there are several limitations in this work that can serve as directions for future research.

With the intention of minimizing background acoustical noise (e.g., sound from air conditioner and/or patient's movements) while providing non-contact measurement, the microphone implemented for signal acquisition is unidirectional and suspended about 0.3 m above the patient's mouth. A drawback of this setup is that patients were requested to sleep in the supine position, and the signal intensity could be degraded when patients turned during sleep. Further research on the optimal placement and polar patterns of microphone in relation to sleep positions may improve the acquisition technique.

The present work studied a limited number of patients (30 apneic and 10 benign); consequently, the training and test data were obtained from the same patient. Future studies should involve a larger sample of subjects, favorably with matched for gender, age, BMI, NC, and AHI, as well as absolutely independent training and test datasets comprising of dissimilar subjects in order to better appraise the diagnostic efficacy of snore-based predictor for OSA. Moreover, for real-time diagnostic applications, one should develop an automatic program that can sequentially execute the following tasks: (a) enhances the acquired signals; (b) detects sleep sound segments; (c) recognizes snores from other sleep sounds; (d) extracts snores exceeding a preset threshold value (e.g., in terms of sound pressure level); (e) analyzes snore properties; and (f) classifies the patient as apneic or benign, together with a predictive AHI value.

The biophysics of snore production is inherently sophisticated due to the involvements of different soft tissues with inhomogeneity compositions and the

dynamic UA anatomical structures. Thus, the analytical results from the synthetic snores require further justification on the methodology for snore synthesis and its relationship with physiological and physical reality. Explicitly, we assumed that the snore ES and the UA are linearly separable, in accordance with the source-filter theory. This theory serves as a basic principle underlying numerous applications in speech technology since a non-interactive source-filter model is simple and the errors introduced by the assumption are approximately negligible in many cases. However, the assumption may not strictly hold for the case of snore sounds. The vibration of soft tissue can be affected by the sound pressure within the UA, entailing a certain degree of coupling between the ES and the UA when the flabby tissue reopens from its shut position. Nonetheless, the coupling may be weakened when the tissue is at its closing position, favoring the independence assumption of the ES and the UA, which makes the processing of snore signals tractable. Future research should consider the effect of source-filter interaction caused by the coupling, model the UA using an autoregressive moving average filter, and further validate the reliability of the proposed OSA diagnostic markers (F1, PF1, PSF, and psychoacoustic metrics including loudness and annoyance).

Apart from the linearity assumption made, we changed the CSA of PX and OC, without taking into account that such changes could lead to modifications of UA dimensions in other anatomical regions. We also stimulated the acoustic waves propagating through the entire UA, but neglected the fact that these waves could be initiated in any locations where there are imbalance of forces between the UA dilator muscles and the negative intraluminal pressure within the UA. Nevertheless, based on the findings from the three SFs and the two clinical UA profiles with different UA lengths, one may reasonably predict that there is an association between the UA

anatomical structures and the acoustical and perceptual characteristics of snores. To further improve naturalness of snore synthesis, one should incorporate the following components into the existing UA acoustic model: (a) acoustic parameters to account for tissue biomechanical properties and UA neuromuscular activities; (b) one or more snore ES circuits to generate SF (e.g., a self-oscillating soft palate circuit); and (c) an acoustic side branch to represent the nasal cavity.

Lastly, owing to limited medical resources, no clinical experiments were conducted to warrant the accuracy of the proposed SFD model in relation to the occurrence and development of physiological events, such as closing, opening, and speed of ES vibration during snoring. Future studies should examine the dynamics of ES by means of sleep nasendoscopy, which is a technique for assessing snoring under light anesthesia, and quantify the SF and SFD pulses, in terms of temporal and spectral dimensions. A variety of SF and SFD waveforms from different snoring modes (e.g., vibrations of soft palate, uvula, tonsils, tongue base, epiglottis, and/or lateral pharyngeal walls) should also be evaluated and compared across multiple acoustic dimensions, thereby providing a greater appreciation of the cause-effect relationships between the physiological or anatomical changes and the acoustical or perceptual characteristics of snores, which will certainly contribute to the development of reliable snore-based screening tools for OSA, the understanding of the pathophysiology of snoring, and the implementation of new therapies for snoring.

# Author's Publications

[A1]  A. K. Ng* and T. S. Koh, "Analysis and modeling of snore source flow with its preliminary application to synthetic snore generation," *IEEE Transactions on Biomedical Engineering*, 2009, doi: 10.1109/TBME.2009.2034139.

[A2]  A. K. Ng*, T. S. Koh, E. Baey, and K. Puvanendran, "Role of upper airway dimensions in snore production: Acoustical and perceptual findings," *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1807-1817, 2009.

[A3]  A. K. Ng*, T. S. Koh, U. R. Abeyratne, and K. Puvanendran, "Investigation of obstructive sleep apnea using nonlinear interactions in nonstationary snore signals," *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1796-1806, 2009.

[A4]  A. K. Ng*, T. S. Koh, K. Puvanendran, and U. R. Abeyratne, "Snore signal enhancement and activity detection via translation-invariant wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 10, pp. 2332-2342, 2008.

[A5]  A. K. Ng*, T. S. Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran, "Could formant frequencies of snore signals be alternative means for the diagnosis of obstructive sleep apnea?," *Sleep Medicine*, vol. 9, no. 8, pp. 894-898, 2008.

[A6]  A. K. Ng*, Z. J. Koh, and T. S. Koh, "Effects of microphone position on snore signal quality and patient comfort," in *Proceedings of the 9th International Conference on Signal Processing*, 2008, pp. 2130-2133.

---

* Asterisk indicates corresponding author.

[A7]   A. K. Ng* and T. S. Koh, "Using psychoacoustics of snoring sounds to screen for obstructive sleep apnea," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 1647-1650.

[A8]   A. K. Ng*, K. Y. Wong, C. H. Tan, and T. S. Koh, "Bispectral analysis of snore signals for obstructive sleep apnea detection," in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 6196-6199.

[A9]   A. K. Ng*, T. S. Koh, L. L. Lim, and K. Puvanendran, "Psychoacoustic evaluation of snoring sounds: A useful supplement to polysomnographic assessment?," *Sleep and Biological Rhythms*, vol. 5, pp. A39-A40, 2007.

[A10]  A. K. Ng*, T. S. Koh, and C. H. Thng, "A level-wavelet-dependent scheme for image denoising via undecimated wavelet transform," in *Proceedings of the 9th IASTED International Conference on Signal and Image Processing*, 2007, pp. 79-83.

[A11]  A. K. Ng*, T. S. Koh, and L. L. Lim, "Snore signals analysis: Review and clinical applications for obstructive sleep apnea," in *Proceedings of the World Congress on Bioengineering*, 2007, pp. 1-4.

[A12]  A. K. Ng* and T. S. Koh, "Acoustic signals enhancement by level-wavelet-dependent thresholding of undecimated wavelets," in *Proceedings of the World Congress on Bioengineering*, 2007, pp. 1-4.

[A13]  A. K. Ng* and T. S. Koh, "A high-fidelity acquisition system for snore signals: Design and implementation," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, 2007, pp. 189-193, art. no. 4258108.

[A14] A. K. Ng*, T. S. Koh, E. Baey, and K. Puvanendran, "Could formant frequencies of snore signals be alternative means for the diagnosis of obstructive sleep apnea?," *Sleep Medicine*, vol. 8, pp. S57, 2007. (Awarded Young Investigator Award at the 2nd World Congress of the World Association of Sleep Medicine in 2007)

[A15] A. K. Ng*, T. S. Koh, E. Baey, and K. Puvanendran, "Speech-like analysis of snore signals for the detection of obstructive sleep apnea," in *Proceedings of the International Conference on Biomedical and Pharmaceutical Engineering*, 2006, pp. 99-103. (Nominated for Outstanding Paper Award)

[A16] A. K. Ng*, T. S. Koh, E. Baey, and K. Puvanendran, "Diagnosis of obstructive sleep apnea using formant features of snore signals," in *IFMBE Proceedings of the World Congress on Medical Physics and Biomedical Engineering*, 2006, pp. 967-970.

# Bibliography

[1]     T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New England Journal of Medicine*, vol. 328, no. 17, pp. 1230-1235, 1993.

[2]     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," *Archives of Internal Medicine*, vol. 162, no. 8, pp. 893-900, 2002.

[3]     K. Puvanendran and K. L. Goh, "From snoring to sleep apnea in a Singapore population," *Sleep Research Online*, vol. 2, no. 1, pp. 11-14, 1999.

[4]     B. Lam, D. C. L. Lam, and M. S. M. Ip, "Obstructive sleep apnoea in Asia," *International Journal of Tuberculosis and Lung Disease*, vol. 11, no. 1, pp. 2-11, 2007.

[5]     M. D. S. Center, "Obstructive sleep apnea," 2009, available: http://www.tiredofcpap.com/Sleep_Apnea.html.

[6]     R. L. Horner, "Pathophysiology of obstructive sleep apnea," *Journal of Cardiopulmonary Rehabilitation and Prevention*, vol. 28, no. 5, pp. 289-298, 2008.

[7]     S. P. Patil, H. Schneider, A. R. Schwartz, and P. L. Smith, "Adult obstructive sleep apnea: Pathophysiology and diagnosis," *Chest*, vol. 132, no. 1, pp. 325-337, 2007.

[8]     K. Banno and M. H. Kryger, "Sleep apnea: Clinical investigations in humans," *Sleep Medicine*, vol. 8, no. 4, pp. 400-426, 2007.

[9]     D. M. Hiestand, P. Britz, M. Goldman, and B. Phillips, "Prevalence of symptoms and risk of sleep apnea in the US population: Results from the National Sleep Foundation Sleep in America 2005 Poll," *Chest*, vol. 130, no. 3, pp. 780-786, 2006.

[10]    K. F. Whyte, M. B. Allen, A. A. Jeffrey, G. A. Gould, and N. J. Douglas, "Clinical features of the sleep apnoea/hypopnoea syndrome," *Quarterly Journal of Medicine*, vol. 72, no. 267, pp. 659-666, 1989.

[11]    A. I. Sanchez and G. Buela-Casal, "Assessment of daytime symptoms in snoring subjects and obstructive sleep apnea patients," *Salud Mental*, vol. 30, no. 1, pp. 9-15, 2007.

[12]    M. Nowak, J. Kornhuber, and R. Meyrer, "Daytime impairment and neurodegeneration in OSAS," *Sleep*, vol. 29, no. 12, pp. 1521-1530, 2006.

[13]    M. P. Accattoli, G. Muzi, M. Dellomo, M. Mazzoli, V. Genovese, G. Palumbo, and G. Abbritti, "Occupational accidents, work performance and obstructive sleep apnea syndrome (OSAS)," *Giornale Italiano di Medicina del Lavoro ed Ergonomia*, vol. 30, no. 3, pp. 297-303, 2008.

[14]    F. Pichel, C. Zamarron, F. Magan, and J. R. Rodriguez, "Sustained attention measurements in obstructive sleep apnea and risk of traffic accidents," *Respiratory Medicine*, vol. 100, no. 6, pp. 1020-1027, 2006.

[15]    T. D. Bradley and J. S. Floras, "Obstructive sleep apnoea and its cardiovascular consequences," *The Lancet*, vol. 373, no. 9657, pp. 82-93, 2009.

[16]    A. Culebras, "Cerebrovascular disease and the pathophysiology of obstructive sleep apnea," *Current Neurology and Neuroscience Reports*, vol. 7, no. 2, pp. 173-179, 2007.

[17]    A. Okcay, V. K. Somers, and S. M. Caples, "Obstructive sleep apnea and hypertension," *Journal of Clinical Hypertension*, vol. 10, no. 7, pp. 549-555, 2008.

[18]    L. Luthje and S. Andreas, "Obstructive sleep apnea and coronary artery disease," *Sleep Medicine Reviews*, vol. 12, no. 1, pp. 19-31, 2008.

[19]    A. H. Anselm, N. Gauthier, R. S. B. Beanlands, and H. Haddad, "Sleep apnea in chronic heart failure," *Current Opinion in Cardiology*, vol. 23, no. 2, pp. 121-126, 2008.

[20]    P. M. Alapat, "Obstructive sleep apnea and stroke," *Current Respiratory Medicine Reviews*, vol. 3, no. 4, pp. 241-244, 2007.

[21]    C. Lopes, A. M. Esteves, L. R. A. Bittencourt, S. Tufik, and M. T. Mello, "Relationship between the quality of life and the severity of obstructive sleep apnea syndrome," *Brazilian Journal of Medical and Biological Research*, vol. 41, no. 10, pp. 908-913, 2008.

[22]    A. Glebocka, A. Kossowska, and M. Bednarek, "Obstructive sleep apnea and the quality of life," *Journal of Physiology and Pharmacology*, vol. 57, no. 4 (suppl.), pp. 111-117, 2006.

[23]    M. M. Ohayon, C. Guilleminault, R. G. Priest, and M. Caulet, "Snoring and breathing pauses during sleep: Telephone interview survey of a United Kingdom population sample," *British Medical Journal*, vol. 314, no. 7084, pp. 860-863, 1997.

[24]    E. Lugaresi, F. Cirignotta, G. Coccagna, and A. Baruzzi, "Snoring and the obstructive apnea syndrome," *Electroencephalography and Clinical Neurophysiology (Supplement)*, no. 35, pp. 421-430, 1982.

[25]    V. Hoffstein, "Snoring," *Chest*, vol. 109, no. 1, pp. 201-222, 1996.

[26]    Y. Hori, H. Shizuku, A. Kondo, H. Nakagawa, B. Kalubi, and N. Takeda, "Endoscopic evaluation of dynamic narrowing of the pharynx by the Bernouilli effect producing maneuver in patients with obstructive sleep apnea syndrome," *Auris Nasus Larynx*, vol. 33, no. 4, pp. 429-432, 2006.

[27]    I. Fajdiga, "Snoring imaging: Could bernoulli explain it all?," *Chest*, vol. 128, no. 2, pp. 896-901, 2005.

[28]    N. Gavriely and O. Jensen, "Theory and measurements of snores," *Journal of Applied Physiology*, vol. 74, no. 6, pp. 2828-2837, 1993.

[29]    E. Olszewska, A. Sieskiewicz, J. Rozycki, M. Rogalewski, E. Tarasow, M. Rogowski, and J. Kulikowska, "A comparison of cephalometric analysis using radiographs and craniofacial computed tomography in patients with obstructive sleep apnea syndrome: Preliminary report," *European Archives of Oto-Rhino-Laryngology*, vol. 266, no. 4, pp. 535-542, 2009.

[30]    B. A. Stuck and J. T. Maurer, "Airway evaluation in obstructive sleep apnea," *Sleep Medicine Reviews*, vol. 12, no. 6, pp. 411-436, 2008.

[31]    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," *Journal of Biomechanics*, vol. 40, no. 10, pp. 2207-2213, 2007.

[32]    J. V. Holland, W. G. Dement, and D. M. Raynall, "Polysomnography: A response to a need for improved communication," in *Proceedings of the 14th Annual Meeting of the Association for the Psychophsyiological Study of Sleep*, 1974.

[33]    B. V. Vaughn and P. Giallanza, "Technical review of polysomnography," *Chest*, vol. 134, no. 6, pp. 1310-1319, 2008.

[34]    M. Reite, J. Ruddy, and K. Nagel, *Concise Guide to Evaluation and Management of Sleep Disorders*. American Psychiatric Press, 2002.

[35]    G. J. Gibson, "Obstructive sleep apnoea syndrome: Underestimated and undertreated," *British Medical Bulletin*, vol. 72, pp. 49-64, 2004.

[36]    T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705-706, 1997.

[37]    S. A. Mickelson, "Preoperative and postoperative management of obstructive sleep apnea patients," *Otolaryngologic Clinics of North America*, vol. 40, no. 4, pp. 877-889, 2007.

[38]    M. M. Harrison, A. Childs, and P. E. Carson, "Incidence of undiagnosed sleep apnea in patients scheduled for elective total joint arthroplasty," *Journal of Arthroplasty*, vol. 18, no. 8, pp. 1044-1047, 2003.

[39]    J. K. Fleischman, R. Ananthamoorthy, H. Greenberg, C. Harvey, and J. P. Merlino, "An unexplained death in the psychiatric emergency room: a case of undiagnosed obstructive sleep apnea?," *General Hospital Psychiatry*, vol. 30, no. 1, pp. 83-86, 2008.

[40]    A. Lofsky, "Sleep apnea and narcotic postoperative pain medication: A morbidity and mortality risk," *Anesthesia Patient Safety Foundation Newsletter*, vol. 17, pp. 24-25, 2002.

[41]    V. Wittmann and D. O. Rodenstein, "Health care costs and the sleep apnea syndrome," *Sleep Medicine Reviews*, vol. 8, no. 4, pp. 269-279, 2004.

[42]    V. Kapur, D. K. Blough, R. E. Sandblom, R. Hert, J. B. De Maine, S. D. Sullivan, and B. M. Psaty, "The medical cost of undiagnosed sleep apnea," *Sleep*, vol. 22, no. 6, pp. 749-755, 1999.

[43]    N. Alghanim, V. R. Comondore, J. Fleetham, C. A. Marra, and N. T. Ayas, "The economic impact of obstructive sleep apnea," *Lung*, vol. 186, no. 1, pp. 7-12, 2008.

[44]    B. Yegneswaran and F. Chung, "The importance of screening for obstructive sleep apnea before surgery," *Sleep Medicine*, vol. 10, no. 2, pp. 270-271, 2009.

[45]    M. Hirshkowitz, "The clinical consequences of obstructive sleep apnea and associated excessive sleepiness," *Journal of Family Practice*, vol. 57, no. 8 (suppl.), 2008.

[46]    C. Planes, M. P. D'Ortho, A. Foucher, M. Berkani, K. Leroux, M. Essalhi, C. Delciaux, M. A. Quera-Salva, and F. Lofaso, "Efficacy and cost of home-initiated auto-nCPAP versus conventional nCPAP," *Sleep*, vol. 26, no. 2, pp. 156-160, 2003.

[47]    L. D. Victor, "Obstructive sleep apnea," *American Family Physician*, vol. 60, no. 8, pp. 2279-2286, 1999.

[48]    J. Rodsutti, M. Hensley, A. Thakkinstian, C. D'Este, and J. Attia, "A clinical decision rule to prioritize polysomnography in patients with suspected sleep apnea," *Sleep*, vol. 27, no. 4, pp. 694-699, 2004.

[49]    A. Sharafkhaneh, P. Richardson, and M. Hirshkowitz, "Sleep apnea in a high risk population: A study of veterans health administration beneficiaries," *Sleep Medicine*, vol. 5, no. 4, pp. 345-350, 2004.

[50]    W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to diagnosis and treatment of patients with suspected sleep apnea,"

*American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 6, pp. 668-672, 2004.

[51]   F. J. Alvarez, I. Fierro, M. T. Gomez-Talegon, A. Vicondoa, and M. Ozcoidi, "Patients treated with obstructive sleep apnea syndrome and fitness to drive assessment in clinical practice in Spain at the medical traffic centers," *Traffic Injury Prevention*, vol. 9, no. 2, pp. 168-172, 2008.

[52]   S. Mazza, J. L. Pepin, B. Naegele, E. Rauch, C. Deschaux, P. Ficheux, and P. Levy, "Driving ability in sleep apnoea patients before and after CPAP treatment: Evaluation on a road safety platform," *European Respiratory Journal*, vol. 28, no. 5, pp. 1020-1028, 2006.

[53]   C. F. P. George, "Reduction in motor vehicle collisions following treatment of sleep apnoea with nasal CPAP," *Thorax*, vol. 56, no. 7, pp. 508-512, 2001.

[54]   V. E. Olivares, "Scheduling strategies," *Radiology Management*, vol. 12, no. 3, pp. 29-30, 1990.

[55]   H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard," *Journal of Sleep Research*, vol. 18, no. 1, pp. 74-84, 2009.

[56]   T. Ogawa, R. Enciso, W. H. Shintaku, and G. T. Clark, "Evaluation of cross-section airway configuration of obstructive sleep apnea," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*, vol. 103, no. 1, pp. 102-108, 2007.

[57] Z. Lan, A. Itoi, M. Takashima, M. Oda, and K. Tomoda, "Difference of pharyngeal morphology and mechanical property between OSAHS patients and normal subjects," *Auris Nasus Larynx*, vol. 33, no. 4, pp. 433-439, 2006.

[58] J. B. Laflen, C. L. Lazarus, and M. R. Amin, "Pitch deviation analysis of pathological voice in connected speech," *Annals of Otology, Rhinology and Laryngology*, vol. 117, no. 2, pp. 90-97, 2008.

[59] D. D. Mehta and R. E. Hillman, "Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods," *Current Opinion in Otolaryngology and Head and Neck Surgery*, vol. 16, no. 3, pp. 211-215, 2008.

[60] K. Shama, A. Krishna, and N. U. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology," *EURASIP Journal on Advances in Signal Processing*, 2007, doi: 10.1155/2007/85286.

[61] Y. Zhang and J. J. Jiang, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," *Journal of Voice*, vol. 22, no. 1, pp. 1-9, 2008.

[62] Y. Zhang and J. J. Jiang, "Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations," *Journal of Sound and Vibration*, vol. 316, pp. 248-262, 2008.

[63] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering Online*, vol. 6, art. no. 23, 2007.

[64] M. Little, P. McSharry, I. Moroz, and S. Roberts, "Nonlinear, biophysically-informed speech pathology detection," in *Proceedings of the IEEE*

*International Conference on Acoustics, Speech and Signal Processing*, 2006, pp. 1080-1083.

[65]   J. H. L. Hansen, L. Gavidia-Ceballos, and J. F. Kaiser, "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 3, pp. 300-313, 1998.

[66]   R. J. Porter and D. M. Hogue, "Nonlinear dynamical systems in speech perception and production," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 2, no. 2, pp. 95-131, 1998.

[67]   H. M. Teager and S. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Speech Production and Speech Modeling*, vol. 55, W. J. Hardcastle and A. Marchal, Eds. London: Kluwer Academic, 1990, pp. 241-262.

[68]   A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Los Angeles: Brain Information Service/Brain Research Institutes, University of California, 1968.

[69]   D. S. Morillo, J. L. Rojas, L. F. Crespo, A. Leon, and N. Gross, "Poincare analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hypopnea syndrome," *Physiological Measurement*, vol. 30, no. 4, pp. 405-420, 2009.

[70]   D. Alvarez, R. Hornero, M. Garcia, F. Del Campo, and C. Zamarron, "Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure," *Artificial Intelligence in Medicine*, vol. 41, no. 1, pp. 13-24, 2007.

[71]    R. Hornero, D. Alvarez, D. Abasolo, F. Del Campo, and C. Zamarron, "Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 107-113, 2007.

[72]    K. P. Pang, C. G. Gourin, and D. J. Terris, "A comparison of polysomnography and the WatchPAT in the diagnosis of obstructive sleep apnea," *Otolaryngology - Head and Neck Surgery*, vol. 137, no. 4, pp. 665-668, 2007.

[73]    G. Weinreich, J. Armitstead, V. Topfer, Y. M. Wang, Y. Wang, and H. Teschler, "Validation of apnealink as screening device for cheyne-stokes respiration," *Sleep*, vol. 32, no. 4, pp. 553-557, 2009.

[74]    S. S. Grover and S. D. Pittman, "Automated detection of sleep disordered breathing using a nasal pressure monitoring device," *Sleep and Breathing*, vol. 12, no. 4, pp. 339-345, 2008.

[75]    J. I. Salisbury and Y. Sun, "Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique," *Medical Engineering and Physics*, vol. 29, no. 3, pp. 336-343, 2007.

[76]    F. R. De Almeida, N. T. Ayas, R. Otsuka, H. Ueda, P. Hamilton, F. C. Ryan, and A. A. Lowe, "Nasal pressure recordings to detect obstructive sleep apnea," *Sleep and Breathing*, vol. 10, no. 2, pp. 62-69, 2006.

[77]    H. M. Al-Angari and A. V. Sahakian, "Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1900-1904, 2007.

[78]  C. Zamarron, R. Hornero, F. Del Campo, D. Abasolo, and D. Alvarez, "Heart rate regularity analysis obtained from pulse oximetric recordings in the diagnosis of obstructive sleep apnea," *Sleep and Breathing*, vol. 10, no. 2, pp. 83-89, 2006.

[79]  A. Mueller, I. Fietze, R. Voelker, S. Eddicks, M. Glos, G. Baumann, and H. Theres, "Screening for sleep-related breathing disorders by transthoracic impedance recording integrated into a Holter ECG system," *Journal of Sleep Research*, vol. 15, no. 4, pp. 455-462, 2006.

[80]  R. Ishida, Y. Yonezawa, H. Maki, H. Ogawa, I. Ninomiya, K. Sada, S. Hamada, A. W. Hahn, and W. M. Caldwell, "A wearable, mobile phone-based respiration monitoring system for sleep apnea syndrome detection," *Biomedical Sciences Instrumentation*, vol. 41, pp. 289-293, 2005.

[81]  S. E. Brietzke and E. A. Mair, "Acoustical analysis of pediatric snoring: What can we learn?," *Otolaryngology - Head and Neck Surgery*, vol. 136, no. 4, pp. 644-648, 2007.

[82]  P. G. Michaelson, P. Allan, J. Chaney, and E. A. Mair, "Validations of a portable home sleep study with twelve-lead polysomnography: Comparisons and insights into a variable gold standard," *Annals of Otology, Rhinology and Laryngology*, vol. 115, no. 11, pp. 802-809, 2006.

[83]  S. Su, F. M. Baroody, M. Kohrman, and D. Suskind, "A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome," *Otolaryngology - Head and Neck Surgery*, vol. 131, no. 6, pp. 844-850, 2004.

[84]  T. Penzel, W. Althaus, K. Meinzer, J. H. Peter, and P. Von Wichert, "A device for ambulatory heart rate, oxygen saturation and snoring recording," in

*Proceedings of the 13th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, 1991, pp. 1616-1617.

[85]   T. Penzel, G. Amend, K. Meinzer, J. H. Peter, and P. Von Wichert, "MESAM: A heart rate and snoring recorder for detection of obstructive sleep apnea," *Sleep*, vol. 13, no. 2, pp. 175-182, 1990.

[86]   F. Cirignotta, S. Mondini, R. Gerardi, B. Mostacci, and E. Sancisi, "Unreliability of automatic scoring of MESAM 4 in assessing patients with complicated obstructive sleep apnea syndrome," *Chest*, vol. 119, no. 5, pp. 1387-1392, 2001.

[87]   M. Koziej, J. K. Cieslicki, K. Gorzelak, P. Sliwinski, and J. Zieliniski, "Hand-scoring of MESAM 4 recordings is more accurate than automatic analysis in screening for obstructive sleep apnoea," *European Respiratory Journal*, vol. 7, no. 10, pp. 1771-1775, 1994.

[88]   R. Stoohs and C. Guilleminault, "MESAM 4: An ambulatory device for the detection of patients at risk for obstructive sleep apnea syndrome (OSAS)," *Chest*, vol. 101, no. 5, pp. 1221-1227, 1992.

[89]   K. Wilson, R. A. Stoohs, T. F. Mulrooney, L. J. Johnson, C. Guilleminault, and Z. Huang, "The snoring spectrum: Acoustic assessment of snoring sound intensity in 1,139 individuals undergoing polysomnography," *Chest*, vol. 115, no. 3, pp. 762-770, 1999.

[90]   T. D. Rossing, F. Dunn, W. M. Hartmann, D. M. Campbell, and N. H. Fletcher, *Springer Handbook of Acoustics*. Springer, 2007.

[91]   D. L. Van Brunt, K. L. Lichstein, S. L. Noe, R. N. Aguillard, and K. W. Lester, "Intensity pattern of snoring sounds as a predictor for sleep-disordered breathing," *Sleep*, vol. 20, no. 12, pp. 1151-1156, 1997.

[92] F. G. Issa, D. Morrison, E. Hadjuk, A. Iyer, T. Feroah, and J. E. Remmers, "Digital monitoring of sleep-disordered breathing using snoring sound and arterial oxygen saturation," *American Review of Respiratory Disease*, vol. 148, no. 4, pp. 1023-1029, 1993.

[93] H. Hara, N. Murakami, Y. Miyauchi, and H. Yamashita, "Acoustic analysis of snoring sounds by a multidimensional voice program," *Laryngoscope*, vol. 116, no. 3, pp. 379-381, 2006.

[94] H. Michael, S. Andreas, B. Thomas, H. Beatrice, H. Werner, and K. Holger, "Analysed snoring sounds correlate to obstructive sleep disordered breathing," *European Archives of Oto-Rhino-Laryngology*, vol. 265, no. 1, pp. 105-113, 2008.

[95] J. R. Perez-Padilla, E. Slawinski, L. M. Difrancesco, R. R. Feige, J. E. Remmers, and W. A. Whitelaw, "Characteristics of the snoring noise in patients with and without occlusive sleep apnea," *American Review of Respiratory Disease*, vol. 147, no. 3, pp. 635-644, 1993.

[96] J. A. Fiz, J. Abad, R. Jane, M. Riera, M. A. Mananas, P. Caminal, D. Rodenstein, and J. Morera, "Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea," *European Respiratory Journal*, vol. 9, no. 11, pp. 2365-2370, 1996.

[97] U. R. Abeyratne, A. S. Wakwella, and C. Hukins, "Pitch jump probability measures for the analysis of snoring sounds in apnea," *Physiological Measurement*, vol. 26, no. 5, pp. 779-798, 2005.

[98] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Silence-breathing-snore classification from snore-related sounds," *Physiological Measurement*, vol. 29, no. 2, pp. 227-243, 2008.

[99]    F. Dalmasso and R. Prota, "Snoring: Analysis, measurement, clinical implications and applications," *European Respiratory Journal*, vol. 9, no. 1, pp. 146-159, 1996.

[100]   R. Beck, M. Odeh, A. Oliven, and N. Gavriely, "The acoustic properties of snores," *European Respiratory Journal*, vol. 8, no. 12, pp. 2120-2128, 1995.

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

[102]   T. H. Lee and U. R. Abeyratne, "Analysis of snoring sounds for the detection of obstructive sleep apnea," *Medical and Biological Engineering and Computing*, vol. 37, no. 2 (suppl.), pp. 538-539, 1999.

[103]   D. Pevernagie, "European guidelines for the accreditation of Sleep Medicine Centres," *Journal of Sleep Research*, vol. 15, no. 2, pp. 231-238, 2006.

[104]   J. Eargle, *The Microphone Book*. Focal Press, 2004.

[105]   P. P. Caffier, J. C. Berl, A. Muggli, A. Reinhardt, A. Jakob, M. Moser, I. Fietze, H. Scherer, and M. Holzl, "Snoring noise pollution - The need for objective quantification of annoyance, regulatory guidelines and mandatory therapy for snoring," *Physiological Measurement*, vol. 28, no. 1, pp. 25-40, 2007.

[106]   M. Herzog, T. Kuhnel, T. Bremert, B. Herzog, W. Hosemann, and H. Kaftan, "The impact of the microphone position on the frequency analysis of snoring sounds," *European Archives of Oto-Rhino-Laryngology*, 2008, doi: 10.1007/s00405-008-0858-7.

[107]    H. T. Friis, "Noise figure of radio receivers," in *Proceedings of the IRE*, 1944, pp. 419-422.

[108]    R. Van Der Togt, E. J. Van Lieshout, R. Hensbroek, E. Beinat, J. M. Binnekade, and P. J. M. Bakker, "Electromagnetic interference from radio frequency identification inducing potentially hazardous incidents in critical care medical equipment," *Journal of the American Medical Association*, vol. 299, no. 24, pp. 2884-2890, 2008.

[109]    S. E. Lapinsky and A. C. Easty, "Electromagnetic interference in critical care," *Journal of Critical Care*, vol. 21, no. 3, pp. 267-270, 2006.

[110]    R. S. Khandpur, *Biomedical Instrumentation: Technology and Applications*. McGraw-Hill Professional, 2004.

[111]    P. Horowitz and W. Hill, *The Art of Electronics*. Cambridge University Press, 1989.

[112]    P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70-73, 1967.

[113]    P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[114]    L. Tan, *Digital Signal Processing: Fundamentals and Applications*. Academic Press, 2007.

[115]    F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51-83, 1978.

[116]    E. A. P. Habets, "Room impulse response generator," 2008, available: http://home.tiscali.nl/ehabets/rir_generator.html.

[117] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943-950, 1979.

[118] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527-1529, 1986.

[119] H. K. Schutte and W. Seidner, "Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/phonetography," *Folia Phoniatrica*, vol. 35, no. 6, pp. 286-288, 1983.

[120] M. S. Keshner, "1/f noise," *Proceedings of the IEEE*, vol. 70, no. 3, pp. 212-218, 1982.

[121] F. Liu, G. Liu, and Z. Zhang, "Representation of 1/f signal with wavelet bases," *Science in China, Series E: Technological Sciences*, vol. 43, no. 4, pp. 380-386, 2000.

[122] G. W. Wornell, "Wavelet-based representations for the 1/f family of fractal processes," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1428-1450, 1993.

[123] Q. Pan, L. Zhang, G. Z. Dai, and H. C. Zhang, "Two denoising methods by wavelet transform," *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3401-3406, 1999.

[124] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, vol. 103, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, pp. 125-150.

[125] E. Pardo, J. L. San Emeterio, M. A. Rodriguez, and A. Ramos, "Noise reduction in ultrasonic NDT using undecimated wavelet transforms," *Ultrasonics*, vol. 44, pp. e1063-e1067, 2006.

[126]    M. J. Shensa, "The discrete wavelet transform: Wedding the atrous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464-2482, 1992.

[127]    G. Beylkin, "On the representation of operators in bases of compactly supported wavelets," *SIAM Journal on Numerical Analysis*, vol. 29, no. 6, pp. 1716-1740, 1992.

[128]    H. T. Guo, "Theory and applications of the shift-invariant, time-varying and undecimated wavelet transforms," Master's thesis, George R. Brown School of Engineering, Rice University, 1995.

[129]    J. Enders, W. Geng, P. Li, M. W. Frazier, and D. J. Scholl, "The shift-invariant discrete wavelet transform and application to speech waveform analysis," *Journal of the Acoustical Society of America*, vol. 117, no. 4 I, pp. 2122-2133, 2005.

[130]    D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[131]    D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425-455, 1994.

[132]    F. Abramovich, T. C. Bailey, and T. Sapatinas, "Wavelet analysis and its statistical applications," *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 49, no. 1, pp. 1-29, 2000.

[133]    D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200-1224, 1995.

[134]   I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 59, no. 2, pp. 319-351, 1997.

[135]   K. Berkner and R. O. J. Wells, "Smoothness estimates for soft-threshold denoising via translation-invariant wavelet transforms," *Applied and Computational Harmonic Analysis*, vol. 12, no. 1, pp. 1-24, 2002.

[136]   I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[137]   I. Daubechies, "Orthogonal bases of compactly supported wavelets," *Communication on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909-996, 1988.

[138]   S. K. Mitra, *Digital Signal Processing*. McGraw-Hill, 2005.

[139]   D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC, 2007.

[140]   J. Peat and B. Barton, *Medical Statistics: A Guide to Data Analysis and Critical Appraisal*. BMJ Publishing Group, 2005.

[141]   J. E. Fowler and B. Pesquet-Popescu, "An overview on wavelets in source coding, communications, and networks," *EURASIP Journal on Image and Video Processing*, 2007, doi: 10.1155/2007/60539.

[142]   S. Cui, Y. Wang, and J. E. Fowler, "Motion estimation and compensation in the redundant-wavelet domain using triangle meshes," *Signal Processing: Image Communication*, vol. 21, no. 7, pp. 586-598, 2006.

[143]   G. Deslauriers and S. Dubuc, "Symmetric iterative interpolation processes," *Constructive Approximation*, vol. 5, no. 1, pp. 49-68, 1989.

[144] N. Saito and G. Beylkin, "Multiresolution representations using the autocorrelation functions of compactly supported wavelets," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3584-3590, 1993.

[145] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells Jr, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10-12, 1996.

[146] U. R. Abeyratne, A. S. Karunajeewa, and C. Hukins, "Mixed-phase modeling in snore sound analysis," *Medical and Biological Engineering and Computing*, vol. 45, no. 8, pp. 791-806, 2007.

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

[148] A. Ouzounov, "A robust feature for speech detection," *Cybernetics and Information Technologies*, vol. 4, no. 2, pp. 3-14, 2004.

[149] Z. S. Liu, X. Y. Luo, H. P. Lee, and C. Lu, "Snoring source identification and snoring noise prediction," *Journal of Biomechanics*, vol. 40, no. 4, pp. 861-870, 2007.

[150] Y. Auregan and C. Depollier, "Snoring: Linear stability analysis and in-vitro experiments," *Journal of Sound and Vibration*, vol. 188, no. 1, pp. 39-53, 1995.

[151] P. C. Loizou, *Speech Enhancement: Theory and Practice*. New York: CRC Press, 2007.

[152] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press, 1999.

[153]   M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag, 1983.

[154]   B. W. Silverman, "Wavelets in statistics: Beyond the standard assumptions," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences (Series A)*, vol. 357, no. 1760, pp. 2459-2473, 1999.

[155]   K. H. Zou, A. J. O'Malley, and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," *Circulation*, vol. 115, no. 5, pp. 654-657, 2007.

[156]   M. Greiner, D. Pfeiffer, and R. D. Smith, "Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests," *Preventive Veterinary Medicine*, vol. 45, no. 1-2, pp. 23-41, 2000.

[157]   M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clinical Chemistry*, vol. 39, no. 4, pp. 561-577, 1993.

[158]   G. Fant, *Speech Acoustics and Phonetics*. Springer, 2006.

[159]   G. Fant, *Acoustic Theory of Speech Production*. The Hague: Mouton, 1960.

[160]   D. G. Childers, *Speech Processing and Synthesis Toolboxes*. John Wiley and Sons, 1999.

[161]   F. J. Owens, *Signal Processing of Speech*. Macmillan, 1993.

[162]   T. H. Lee, U. R. Abeyratne, K. Puvanendran, and K. L. Goh, "Formant-structure and phase-coupling analysis of human snoring sounds for the detection of obstructive sleep apnea," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 3, pp. 243-248, 2001.

[163]   P. P. Vaidyanathan, *The Theory of Linear Prediction*. Morgan and Claypool Publishers, 2008.

[164]   P. E. Papamichalis, *Practical Approaches to Speech Coding*. Prentice Hall, 1987.

[165]   J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561-580, 1975.

[166]   J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. Springer-Verlag, 1976.

[167]   S. Haykin and B. Widrow, *Least-Mean-Square Adaptive Filters*. Wiley-Interscience, 2003.

[168]   G. Walker, "On periodicity in series of related terms," *Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences*, vol. 131, pp. 518-532, 1931.

[169]   G. U. Yule, "On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London, Series A: Physical Sciences and Engineering*, vol. 226, pp. 267-298, 1927.

[170]   U. Grenander, *Toeplitz Forms and Their Applications*. Chelsea Publishing Company, 1984.

[171]   J. Durbin, "The fitting of time series models," *Review of the International Statistical Institute*, vol. 28, no. 3, pp. 233-243, 1960.

[172]   J. Makhoul, "Spectral linear prediction: Properities and applications," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 23, no. 3, pp. 283-296, 1975.

[173] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 411-423, 1991.

[174] R. J. McAulay, "Maximum likelihood spectral estimation and its application to narrowband speech coding," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-32, no. 2, pp. 243-251, 1984.

[175] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electronics and Communications in Japan*, vol. 53A, pp. 36-43, 1970.

[176] C. Kim, K. D. Seo, and W. Sung, "A robust formant extraction algorithm combining spectral peak picking and root polishing," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1-16, 2006.

[177] S. S. McCandless, "An algorithm for automatic formant extraction using linear prediction spectra," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 22, no. 2, pp. 135-141, 1974.

[178] A. Gilat and V. Subramaniam, *Numerical Methods for Engineers and Scientists: An Introduction with Applications Using MATLAB*. Wiley, 2007.

[179] Y.-S. Hsiao and D. G. Childers, "New approach to formant estimation and modification based on pole interaction," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, 1997, pp. 783-787.

[180] J. D. Markel, "Digital inverse filtering - A new tool for formant trajectory estimation," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-20, no. 3, pp. 129-137, 1972.

[181] D. L. Streiner, "Speaking graphically: An introduction to some newer graphing techniques," *Canadian Journal of Psychiatry*, vol. 42, no. 4, pp. 388-394, 1997.

[182] M. Frigge, D. C. Hoaglin, and B. Iglewicz, "Some implementations of the boxplot," *The American Statistician*, vol. 43, no. 1, pp. 50-54, 1989.

[183] V. Hoffstein, S. Mateika, and S. Nash, "Comparing perceptions and measurements of snoring," *Sleep*, vol. 19, no. 10, pp. 783-789, 1996.

[184] F. Gustafsson, "Determining the initial states in forward-backward filtering," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 988-992, 1996.

[185] K. T. Al-Sarayreh, R. E. Al-Qutaish, and B. M. Al-Kasasbeh, "Using the sound recognition techniques to reduce the electricity consumption in highways," *Journal of American Science*, vol. 5, no. 2, pp. 1-12, 2009.

[186] M. Fezari, "Design of speech control system based on HMM approach applied to a robot arm," *International Journal of Soft Computing*, vol. 3, no. 2, pp. 106-111, 2008.

[187] A. O. Afolabi, A. Williams, and O. Dotun, "Development of a text dependent speaker-identification security system," *Research Journal of Applied Sciences*, vol. 2, no. 6, pp. 677-684, 2007.

[188] O. Scharenborg, V. Wan, and R. K. Moore, "Towards capturing fine phonetic variation in speech using articulatory features," *Speech Communication*, vol. 49, no. 10-11, pp. 811-826, 2007.

[189] B. Yegnanarayana and P. Satyanarayana Murthy, "Source-system windowing for speech analysis and synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 2, pp. 133-137, 1996.

[190] S. Park, C. Kim, and I. Jung, "Metric study of upper airway and trachea in nomal korean adults using fiberoptic bronchoscopy: Study of endotracheal tube fixation positioning in adults," *Korean Journal of Anesthesiology*, vol. 31, no. 6, pp. 733-738, 1996.

[191] N. Gavriely and M. Herzberg, "Parametric representation of normal breath sounds," *Journal of Applied Physiology*, vol. 73, no. 5, pp. 1776-1784, 1992.

[192] L. Ye, G. W. Pien, and T. E. Weaver, "Gender differences in the clinical manifestation of obstructive sleep apnea," *Sleep Medicine*, 2009, doi: 10.1016/j.sleep.2009.02.006.

[193] C. M. Lin, T. M. Davidson, and S. Ancoli-Israel, "Gender differences in obstructive sleep apnea and treatment implications," *Sleep Medicine Reviews*, vol. 12, no. 6, pp. 481-496, 2008.

[194] A. S. Jordan, A. Wellman, J. K. Edwards, K. Schory, L. Dover, M. MacDonald, S. R. Patel, R. B. Fogel, A. Malhotra, and D. P. White, "Respiratory control stability and upper airway collapsibility in men and women with obstructive sleep apnea," *Journal of Applied Physiology*, vol. 99, no. 5, pp. 2020-2027, 2005.

[195] V. Mohsenin, "Gender differences in the expression of sleep-disordered breathing: Role of upper airway dimensions," *Chest*, vol. 120, no. 5, pp. 1442-1447, 2001.

[196] U. Anttalainen, T. Saaresranta, N. Kalleinen, J. Aittokallio, T. Vahlberg, and O. Polo, "Gender differences in age and BMI distributions in partial upper airway obstruction during sleep," *Respiratory Physiology and Neurobiology*, vol. 159, no. 2, pp. 219-226, 2007.

[197] G. Bertino, E. Matti, S. Migliazzi, F. Pagella, C. Tinelli, and M. Benazzo, "Acoustic changes in voice after surgery for snoring: preliminary results," *Acta Otorhinolaryngologica Italica*, vol. 26, no. 2, pp. 110-114, 2006.

[198]   A. Behrman, M. J. Shikowitz, and S. Dailey, "The effect of upper airway surgery on voice," *Otolaryngology - Head and Neck Surgery*, vol. 127, no. 1, pp. 36-42, 2002.

[199]   T. Murry and R. C. Bone, "Acoustic characteristics of speech following uvulopalatopharyngoplasty," *Laryngoscope*, vol. 99, no. 12, pp. 1217-1219, 1989.

[200]   M. Herzog, E. Schieb, T. Bremert, B. Herzog, W. Hosemann, H. Kaftan, and T. Kuhnel, "Frequency analysis of snoring sounds during simulated and nocturnal snoring," *European Archives of Oto-Rhino-Laryngology*, vol. 265, no. 12, pp. 1553-1562, 2008.

[201]   S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. E. Camilleri, "Sound frequency analysis and the site of snoring in natural and induced sleep," *Clinical Otolaryngology and Allied Sciences*, vol. 27, no. 3, pp. 162-166, 2002.

[202]   S. Miyazaki, Y. Itasaka, K. Ishikawa, and K. Togawa, "Acoustic analysis of snoring and the site of airway obstruction in sleep related respiratory disorders," *Acta Oto-Laryngologica (Supplement)*, no. 537, pp. 47-51, 1998.

[203]   S. J. Quinn, L. Huang, P. D. M. Ellis, and J. E. F. Williams, "The differentiation of snoring mechanisms using sound analysis," *Clinical Otolaryngology and Allied Sciences*, vol. 21, no. 2, pp. 119-123, 1996.

[204]   T. P. Ryan, *Modern Regression Methods*. Wiley, 2009.

[205]   D. G. Kleinbaum, L. L. Kupper, A. Nizam, and K. E. Muller, *Applied Regression Analysis and Multivariable Methods*. Duxbury Press, 2007.

[206]   S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*. Wiley-Interscience, 2006.

[207]   H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. USA: Oxford University Press, 2004.

[208]   R. M. Howell, A. D. Lucey, P. W. Carpenter, and M. W. Pitman, "Interaction between a cantilevered-free flexible plate and ideal flow," *Journal of Fluids and Structures*, vol. 25, no. 3, pp. 544-566, 2009.

[209]   F. Chouly, A. Van Hirtum, P. Y. Lagree, X. Pelorson, and Y. Payan, "Modelling the human pharyngeal airway: Validation of numerical simulations using in vitro experiments," *Medical and Biological Engineering and Computing*, vol. 47, no. 1, pp. 49-58, 2009.

[210]   F. Chouly, A. Van Hirtum, P. Y. Lagree, X. Pelorson, and Y. Payan, "Numerical and experimental study of expiratory flow in the case of major upper airway obstructions with fluid-structure interaction," *Journal of Fluids and Structures*, vol. 24, no. 2, pp. 250-269, 2008.

[211]   X. Sun, C. Yu, Y. Wang, and Y. Liu, "Numerical simulation of soft palate movement and airflow in human upper airway by fluid-structure interaction method," *Acta Mechanica Sinica/Lixue Xuebao*, vol. 23, no. 4, pp. 359-367, 2007.

[212]   Y. Huang, A. Malhotra, and D. P. White, "Computational simulation of human upper airway collapse using a pressure-/state-dependent model of genioglossal muscle contraction under laminar flow conditions," *Journal of Applied Physiology*, vol. 99, no. 3, pp. 1138-1148, 2005.

[213]   T. Aittokallio, M. Gyllenberg, and O. Polo, "A model of a snorer's upper airway," *Mathematical Biosciences*, vol. 170, no. 1, pp. 79-90, 2001.

[214]  L. Huang and J. E. F. Williams, "Neuromechanical interaction in human snoring and upper airway obstruction," *Journal of Applied Physiology*, vol. 86, no. 6, pp. 1759-1763, 1999.

[215]  L. Huang, S. J. Quinn, P. D. M. Ellis, and J. E. F. Williams, "Biomechanics of snoring," *Endeavour*, vol. 19, no. 3, pp. 96-100, 1995.

[216]  L. Huang, "Mechanical modeling of palatal snoring," *Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3642-3648, 1995.

[217]  J. Soraghan, A. Hussain, A. Alkulabi, and T. S. Durrani, "Higher-order statistics-based nonlinear speech analysis," *Control and Intelligent Systems*, vol. 30, no. 1, pp. 11-18, 2002.

[218]  J. W. A. Fackrell and S. McLaughlin, "The higher-order statistics of speech signals," in *Proceedings of the IEE Colloquium on Techniques for Speech Processing and their Application*, 1994, pp. 7/1-7/6.

[219]  Y. Larsen, A. Hanssen, and H. L. Pecseli, "Analysis of non-stationary mode coupling by means of wavelet-bicoherence," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2001, pp. 3581-3584.

[220]  Y. Larsen and A. Hanssen, "Wavelet-polyspectra: Analysis of non-stationary and non-Gaussian/non-linear signals," in *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing*, 2000, pp. 539-543.

[221]  S. Mallet, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[222]  B. P. Van Milligen, E. Sanctiez, T. Estrada, C. Hidalgo, B. Branas, B. Carreras, and L. Garcia, "Wavelet bicoherence: A new turbulence analysis tool," *Physics of Plasmas*, vol. 2, no. 8, pp. 3017-3032, 1995.

[223] G. Dong, Y. Ma, M. Perlin, X. Ma, B. Yu, and J. Xu, "Experimental study of wave-wave nonlinear interactions using the wavelet-based bicoherence," *Coastal Engineering*, vol. 55, no. 9, pp. 741-752, 2008.

[224] M. A. K. Elsayed, "Use of wavelet bicoherence in analyzing nonlinear wind-wave interaction during wave growth," *Journal of Coastal Research*, vol. 23, no. 6, pp. 1593-1601, 2007.

[225] M. A. K. Elsayed, "Wavelet bicoherence analysis of wind-wave interaction," *Ocean Engineering*, vol. 33, no. 3-4, pp. 458-470, 2006.

[226] A. A. Koronovskii and A. E. Khramov, "Wavelet bicoherence analysis as a method for investigating coherent structures in an electron beam with an overcritical current," *Plasma Physics Reports*, vol. 28, no. 8, pp. 666-681, 2002.

[227] T. Dudok De Wit and V. V. Krasnosel'skikh, "Wavelet bicoherence analysis of strong plasma turbulence at the Earth's quasiparallel bow shock," *Physics of Plasmas*, vol. 2, no. 11, pp. 4307-4311, 1995.

[228] S. A. Taplidou and L. J. Hadjileontiadis, "Nonlinear analysis of wheezes using wavelet bicoherence," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 563-570, 2007.

[229] J. Chung and E. J. Powers, "The statistics of wavelet-based bicoherence," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1998, pp. 141-144.

[230] M. J. Hinich and H. Messer, "On the principal domain of the discrete bispectrum of a stationary signal," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2130-2134, 1995.

[231]  A. Aldroubi and M. Unser, *Wavelets in Medicine and Biology*. CRC Press, 1996.

[232]  H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*. Springer, 2007.

[233]  H. Yang, D. Huang, and L. Cai, "Perceptually weighted mel-cepstrum analysis of speech based on psychoacoustic model," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 12, pp. 2998-3001, 2006.

[234]  J. Zhang, L. Gao, and D. Zhang, "A high-performance psychoacoustics approach to speech quality evaluation," *Information Technology Journal*, vol. 5, no. 3, pp. 485-488, 2006.

[235]  V. Iliadou and S. Kaprinis, "Clinical psychoacoustic in Alzheimer's disease central auditory processing disorders and speech deterioration," *Annals of General Hospital Psychiatry*, vol. 2, art. no. 12, 2003.

[236]  L. N. Kelchner, S. B. Brehm, B. Weinrich, J. Middendorf, A. deAlarcon, L. Levin, and R. Elluru, "Perceptual evaluation of severe pediatric voice disorders: Rater reliability using the consensus auditory perceptual evaluation of voice," *Journal of Voice*, 2008, doi: 10.1016/j.jvoice.2008.09.004.

[237]  R. Linder, A. E. Albers, M. Hess, S. J. Poppl, and R. Schonweiler, "Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features," *Journal of Voice*, vol. 22, no. 2, pp. 155-163, 2008.

[238]  N. D. Cook and T. Hayashi, "The psychoacoustics of harmony perception," *American Scientist*, vol. 96, no. 4, pp. 311-319, 2008.

[239]  S. A. Iakovides, V. T. H. Iliadou, V. T. H. Bizeli, S. G. Kaprinis, K. N. Fountoulakis, and G. S. Kaprinis, "Psychophysiology and psychoacoustics of

music: Perception of complex sound in normal subjects and psychiatric patients," *Annals of General Hospital Psychiatry*, vol. 3, art. no. 6, 2004.

[240] L. P. R. De Oliveira, K. Janssens, P. Gajdatsy, H. Van Der Auweraer, P. S. Varoto, P. Sas, and W. Desmet, "Active sound quality control of engine induced cavity noise," *Mechanical Systems and Signal Processing*, vol. 23, no. 2, pp. 476-488, 2009.

[241] A. Nykanen and A. Sirkka, "Specification of component sound quality applied to automobile power windows," *Applied Acoustics*, vol. 70, no. 6, pp. 813-820, 2009.

[242] Part, J. Yuan, Q. Huang, Z. Liu, L. Hu, and R. Wang, "An application of psychoacoustic metrics to improve sound quality of fission air-conditioning unit," *Lecture Notes in Artifical Intelligence*, vol. 5315, no. 2, pp. 1183-1191, 2008.

[243] S. K. Lee, "Objective evaluation of interior sound quality in passenger cars during acceleration," *Journal of Sound and Vibration*, vol. 310, no. 1-2, pp. 149-168, 2008.

[244] S. Choi, W. Moon, S. Gang, and T. Hwang, "Human-based evaluation of sound from hard disk drives for noise control," *Microsystem Technologies*, vol. 10, no. 8-9, pp. 640-648, 2004.

[245] N. Otto, S. Amman, C. Eaton, and S. Lake, "Guidelines for jury evaluations of automotive sounds," *Sound and Vibration*, vol. 35, no. 4, pp. 24-47, 2001.

[246] P. Virkkula, A. Bachour, M. Hytonen, H. Malmberg, T. Salmi, and P. Maasilta, "Patient- and bed partner-reported symptoms, smoking, and nasal resistance in sleep-disordered breathing," *Chest*, vol. 128, no. 4, pp. 2176-2182, 2005.

[247]   W. J. Strawbridge, S. J. Shema, and R. E. Roberts, "Impact of spouses' sleep problems on partners," *Sleep*, vol. 27, no. 3, pp. 527-531, 2004.

[248]   J. Ulfberg, N. Carter, M. Talback, and C. Edling, "Adverse health effects among women living with heavy snorers," *Health Care for Woman International*, vol. 21, no. 2, pp. 81-90, 2000.

[249]   W. M. Troxel, T. F. Robles, M. Hall, and D. J. Buysse, "Marital quality and the marital bed: Examining the covariation between relationship quality and sleep," *Sleep Medicine Reviews*, vol. 11, no. 5, pp. 389-404, 2007.

[250]   E. O. Pelausa and L. M. Tarshis, "Surgery for snoring," *Laryngoscope*, vol. 99, no. 10 I, pp. 1006-1010, 1989.

[251]   J. R. Dille, "Snoring can be fatal for your marriage and for you," *Aviation Space and Environmental Medicine*, vol. 58, no. 12, pp. 1234, 1987.

[252]   Y. H. Lee, A. Johan, K. K. H. Wong, N. Edwards, and C. Sullivan, "Prevalence and risk factors for obstructive sleep apnea in a multiethnic population of patients presenting for bariatric surgery in Singapore," *Sleep Medicine*, vol. 10, no. 2, pp. 226-232, 2009.

[253]   A. Onat, G. Hergenc, H. Yuksel, G. Can, E. Ayhan, Z. Kaya, and D. Dursunoglu, "Neck circumference as a measure of central obesity: Associations with metabolic syndrome and obstructive sleep apnea syndrome beyond waist circumference," *Clinical Nutrition*, vol. 28, no. 1, pp. 46-51, 2009.

[254]   R. Plywaczewski, P. Bielen, M. Bednarek, L. Jonczak, D. Gorecka, and P. Sliwinski, "Influence of neck circumference and body mass index on obstructive sleep apnoea severity in males," *Pneumonologia i Alergologia Polska*, vol. 76, no. 5, pp. 313-320, 2008.

[255]   C. D. Bertram, "Fluid flow in distensible vessels," *Clinical and Experimental Pharmacology and Physiology*, vol. 36, no. 2, pp. 206-216, 2009.

[256]   C. D. Bertram, "Flow-induced oscillation of collapsed tubes and airway structures," *Respiratory Physiology and Neurobiology*, vol. 163, no. 1-3, pp. 256-265, 2008.

[257]   T. S. Balint and A. D. Lucey, "Instability of a cantilevered flexible plate in viscous channel flow," *Journal of Fluids and Structures*, vol. 20, no. 7, pp. 893-912, 2005.

[258]   M. Airas, "TKK Aparat: An environment for voice inverse filtering and parameterization," *Logopedics Phoniatrics Vocology*, vol. 33, no. 1, pp. 49-64, 2008.

[259]   T. Waaramaa, A. M. Laukkanen, M. Airas, and P. Alku, "Perception of emotional valences and activity levels from vowel segments of continuous speech," *Journal of Voice*, 2008, doi: 10.1016/j.jvoice.2008.04.004.

[260]   L. Lehto, M. Airas, E. Bjorkner, J. Sundberg, and P. Alku, "Comparison of two inverse filtering methods in parameterization of the glottal closing phase characteristics in different phonation types," *Journal of Voice*, vol. 21, no. 2, pp. 138-150, 2007.

[261]   L. Ljung, *System Identification: Theory for the User*. Prentice Hall PTR, 1998.

[262]   H. Akaike, "Power spectrum estimation through autoregressive model fitting," *Annals of the Institute of Statistical Mathematics*, vol. 21, pp. 407-419, 1969.

[263]   H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 719-723, 1974.

[264]   J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Annals of Statistics*, vol. 11, no. 2, pp. 416-431, 1983.

[265]  Z. Zhang and C. Y. Espy-Wilson, "A vocal-tract model of American English /l/," *Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1274-1280, 2004.

[266]  P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, "An acoustic model of the respiratory tract," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 543-550, 2001.

[267]  A. De Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917-1930, 2002.

[268]  D. G. Jung, H. Y. Cho, R. R. Grunstein, and B. Yee, "Predictive value of kushida index and acoustic pharyngometry for the evaluation of upper airway in subjects with or without obstructive sleep apnea," *Journal of Korean Medical Science*, vol. 19, no. 5, pp. 662-667, 2004.

[269]  V. Mohsenin, "Effects of gender on upper airway collapsibility and severity of obstructive sleep apnea," *Sleep Medicine*, vol. 4, no. 6, pp. 523-529, 2003.

[270]  D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394-2410, 1991.

[271]  Y. Qiu, X. Ding, J. Feng, and Z. Mo, "QRS complexes detection based on Mexican-hat wavelet," *Journal of Biomedical Engineering*, vol. 23, no. 6, pp. 1347-1349, 2006.

[272]  Z. Zhou, Y. Du, G. G. Rodney, and M. F. Schneider, "CA2+ sparks detection and classification using Gaussian-Mexican hat wavelet," in *Proceedings of the International Conference on Image Processing*, 2006, art. no. 4379569.

[273] X. Kang, C. J. Tay, C. Quan, and X. Y. He, "Dynamic characterization of MEMS structures by ultrasonic wave excitation," *Journal of Micromechanics and Microengineering*, vol. 17, no. 12, pp. 2426-2431, 2007.

[274] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," in *Speech Transmission Laboratory Quarterly Progress and Status Reports*, 1985, pp. 1-13.

[275] P. Ladefoged, *A Course in Phonetics*. Wadsworth Publishing, 2005.

[276] I. Koutsourelakis, E. Vagiakis, C. Roussos, S. Zakynthinos, and C. Roussos, "Obstructive sleep apnoea and oral breathing in patients free of nasal obstruction," *European Respiratory Journal*, vol. 28, no. 6, pp. 1222-1228, 2006.

[277] K. Gleeson, C. W. Zwillich, K. Braier, and D. P. White, "Breathing route during sleep," *American Review of Respiratory Disease*, vol. 134, no. 1, pp. 115-120, 1986.

[278] D. Sciamarella and G. Artana, "A water hammer analysis of pressure and flow in the voice production system," *Speech Communication*, vol. 51, no. 4, pp. 344-351, 2009.

[279] M. J. Birch and P. D. Srodon, "Biomechanical properties of the human soft palate," *Cleft Palate-Craniofacial Journal*, vol. 46, no. 3, pp. 268-274, 2009.

[280] J. Z. Wu, R. G. Dong, and A. W. Schopper, "Analysis of effects of friction on the deformation behavior of soft tissues in unconfined compression tests," *Journal of Biomechanics*, vol. 37, no. 1, pp. 147-155, 2004.

[281] S. M. Kim, T. M. McCulloch, and K. Rim, "Comparison of viscoelastic properties of the pharyngeal tissue: Human and canine," *Dysphagia*, vol. 14, no. 1, pp. 8-16, 1999.

[282]    Y. Payan, G. Bettega, and B. Raphae, "A biomechanical model of the human tongue and its clinical implication," in *Proceedings of the Conference on Medical Imaging Computing and Computer-Assisted Intervention*, 1998, pp. 688-695.

[283]    Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissues*. Springer, 1993.

[284]    M. Veldi, V. Vasar, A. Vain, and M. Kull, "Obstructive sleep apnea and ageing: Myotonometry demonstrates changes in the soft palate and tongue while awake," *Pathophysiology*, vol. 11, no. 3, pp. 159-165, 2004.

[285]    M. Veldi, V. Vasar, T. Hion, M. Kull, and A. Vain, "Ageing, soft-palate tone and sleep-related breathing disorders," *Clinical Physiology*, vol. 21, no. 3, pp. 358-364, 2001.

[286]    M. Veldi, V. Vasar, T. Hion, A. Vain, and M. Kull, "Myotonometry demonstrates changes of lingual musculature in obstructive sleep apnoea," *European Archives of Oto-Rhino-Laryngology*, vol. 259, no. 2, pp. 108-112, 2002.

[287]    M. Veldi, V. Vasar, A. Vain, T. Hion, and M. Kull, "Computerized endopharyngeal myotonometry (CEM): A new method to evaluate the tissue tone of the soft palate in patients with obstructive sleep apnoea syndrome," *Journal of Sleep Research*, vol. 9, no. 3, pp. 279-284, 2000.

[288]    C. O'Connor-Reina, M. T. Garcia-Iriarte, D. Gomez Angel, and A. Rodriguez-Diaz, "Bipolar radiofrequency uvulopalatoplasty combined with injection snoreplasty: A reasonable option for the problem of snoring," *Journal for Oto-Rhino-Laryngology, Head and Neck Surgery*, vol. 71, no. 2, pp. 105-111, 2009.

[289]   S. E. Brietzke and E. A. Mair, "Injection snoreplasty: Extended follow-up and new objective data," *Otolaryngology - Head and Neck Surgery*, vol. 128, no. 5, pp. 605-615, 2003.

[290]   A. A. Maheshwar, K. G. Gomez, M. Obilanade, and R. A. Evans, "Efficacy of laser palatoplasty: Four-year results," *International Journal of Clinical Practice*, vol. 56, no. 7, pp. 501-503, 2002.

[291]   M. J. Drinnan, H. C. Richardson, P. R. Close, A. J. Smithson, J. E. S. White, C. J. Griffiths, H. F. Marshall, and G. J. Gibson, "Objective benefit of laser palatoplasty for non-apnoeic snoring," *Clinical Otolaryngology and Allied Sciences*, vol. 24, no. 4, pp. 335-338, 1999.

[292]   M. Friedman, P. Schalch, H. C. Lin, K. A. Kakodkar, N. J. Joseph, and N. Mazloom, "Palatal implants for the treatment of snoring and obstructive sleep apnea/hypopnea syndrome," *Otolaryngology - Head and Neck Surgery*, vol. 138, no. 2, pp. 209-216, 2008.

[293]   J. H. Romanow and P. J. Catalano, "Initial U.S. pilot study: Palatal implants for the treatment of snoring," *Otolaryngology - Head and Neck Surgery*, vol. 134, no. 4, pp. 551-557, 2006.

## Appendix A

# Polysomnographic Report of an Apneic Patient

| | |
|---|---|
| Patient's name: | Sleep study number: |
| Identification card number: | Date of study: |
| Date of birth: | Gender: Male |
| Height: 1.85 m | Body mass index: 27.8 kg/m$^2$ |
| Weight: 95 kg | Neck circumference: 43.0 cm |
| Referring physician: | Recording technologist: |

**Study Performed**: Nocturnal polysomnography

**Procedure**: Complete polysomnography with a digital sleep system using the international 10-20 electrode placement for recording electroencephalogram, electroocculogram, electromyogram from chin, electrocardiogram, respiratory effort, oximetry, body position, airflow, snoring sound, pulse rate, and limb movement.

**Summary**: Sleep study was performed from 23:01:30 to 06:51:30. Patient slept at 00:20:30 and woke up at 06:51:30. He mostly slept in the supine position. For both the lateral and supine positions, loud snoring was heard.

**Sleep architecture**: The patient sleep latency was 79 min, and rapid eye movement (REM) latency was 279 min. Sleep period time was 370 min, and total sleep time was

266 min. Sleep efficiency was 56.6% with  24.1% of stage 1 sleep, 64.7% of stage 2 sleep, 10.2% of stage 3 sleep, 0.0% stage 4 sleep, and 1.1% REM. The total number of arousals was 279, and total arousal index was 62.9 per hour. The spontaneous arousal index was 6.3 per hour, and the respiratory arousal index was 56.6 per hour.

**Respiratory disorders**: Respiratory analysis demonstrated 271 obstructive apneas, 0 mixed apneas, and 0 central apneas, with a total of 271 apneas. The apnea index was 61.1 events/h, and apnea-hypopnea index (AHI) was 71.3 events/h. The patient spent 76.5% of total sleep time in the supine position with a supine AHI of 89.9 events/h. The REM AHI was 100.0 events/h versus non-REM AHI of 71.0 events/h.

**Oxyhemoglobin desaturation**: Number of desaturations was 276 with a nadir reaching 81% in stage 2 sleep, followed by apnea, and longest apnea is 42 s occurring in stage 2 sleep.

**Periodic limb movement**: No periodic limb movement was noted.

**Diagnosis**: Patient has a severe obstructive sleep apnea (AHI = 71.3 events/h).

**Sleep staging summary**

| | | | |
|---|---|---|---|
| Recording start time | 23:01:30 | Recording end time | 06:51:30 |
| Analysis start time | 00:20:30 | Analysis end time | 06:30:29 |
| Total recording time (min) | 470 | Epoch size (sec) | 30 |
| Number of awakenings | 82 | Total sleep time (min) | 266 |
| Wake time after sleep onset (min) | 104 | Sleep efficiency (%) | 56.6 |
| Sleep onset latency (min) | 79 | Stage REM latency (min) | 279 |
| Total sleep period time (min) | 370 | | |

**Respiratory event summary**

| Parameter | REM | Non-REM | Sleep |
|---|---|---|---|
| Apnea | 5 | 266 | 271 |
| Hypopnea | 0 | 45 | 45 |
| Apnea and hypopnea | 5 | 311 | 316 |
| Duration of apnea (min) | 1.67 | 108.20 | 109.88 |
| Duration of hypopnea (min) | 0.00 | 17.75 | 17.75 |
| Duration of apnea and hypopnea (min) | 1.67 | 125.97 | 127.65 |
| Apnea-hypopnea index (events/h) | 100.0 | 71.0 | 71.3 |
| Respiratory arousal index (events/h) | 0.0 | 6.4 | 56.6 |

**Heart rate summary**

| | |
|---|---|
| Average heart rate (beats/min) | 58 |
| Slowest heart rate (beats/min) | 37 |
| Fastest heart rate (beats/min) | 83 |

**Oxyhemoglobin summary**

| Parameter | REM | Non-REM | Sleep |
|---|---|---|---|
| Lowest oxyhemoglobin saturation (%) | 84 | 81 | 81 |
| Mean minimum oxyhemoglobin saturation (%) | 92 | 89 | 91 |
| Mean oxyhemoglobin desaturation (%) | - | - | 10 |

**Arousal summary**

| Type of arousal | REM | Non-REM | Sleep |
|---|---|---|---|
| Spontaneous | 0 | 28 | 28 |
| Respiratory | 3 | 248 | 251 |
| Periodic limb movement | 0 | 0 | 0 |
| Total number of arousals | - | - | 279 |
| Spontaneous (per hour) | 0.0 | 6.4 | 6.3 |
| Respiratory (per hour) | 60.0 | 56.6 | 56.6 |
| Periodic limb movement (per hour) | 0.0 | 0.0 | 0.0 |
| Total number of arousals per hour | - | - | 62.9 |

**Graphical summary of polysomnography (Appendix A)**
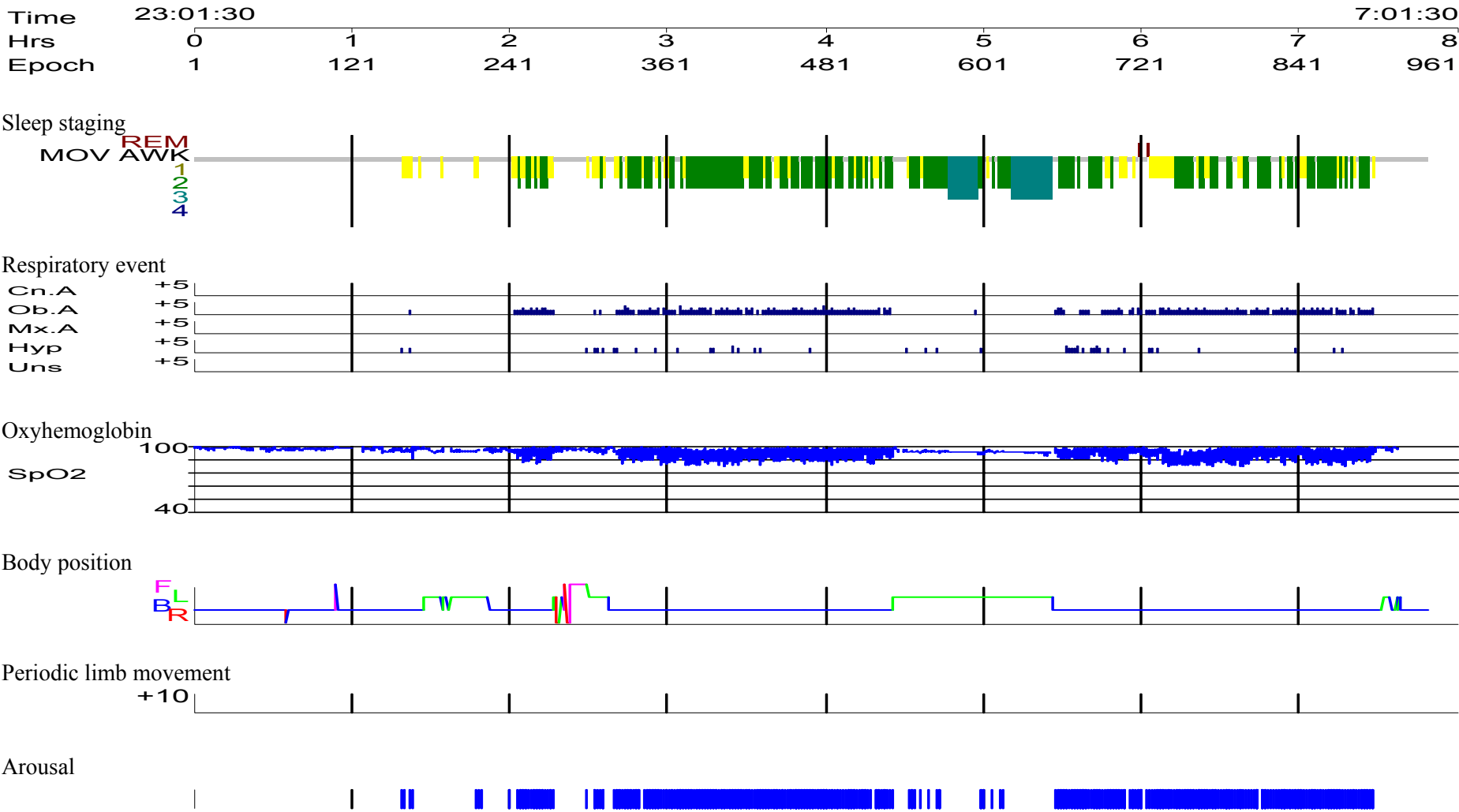
## Appendix B

# Polysomnographic Report of a Benign Patient

Patient's name:                          Sleep study number:

Identification card number:              Date of study:

Date of birth:                           Gender: Male

Height: 1.70 m                           Body mass index: 22.1 kg/m$^2$

Weight: 64 kg                            Neck circumference: 36.5 cm

Referring physician:                     Recording technologist:

**Study Performed**: Nocturnal polysomnography

**Procedure**: Complete polysomnography with a digital sleep system using the international 10-20 electrode placement for recording electroencephalogram, electroocculogram, electromyogram from chin, electrocardiogram, respiratory effort, oximetry, body position, airflow, snoring sound, pulse rate, and limb movement.

**Summary**: Sleep study was performed from 22:50:20 to 05:45:20. Patient slept at 22:57:20 and woke up at 05:45:20. He mostly slept in the lateral position. In the lateral position, no snoring was heard. In the supine position, mild snoring was heard.

**Sleep architecture**: The patient sleep latency was 7 min, and rapid eye movement (REM) latency was 19 min. Sleep period time was 355 min, and total sleep time was

322 min. Sleep efficiency was 77.6% with 7.9% of stage 1 sleep, 48.0% of stage 2 sleep, 6.2% of stage 3 sleep, 15.5% stage 4 sleep, and 22.4% REM. The total number of arousals was 77, and total arousal index was 14.3 per hour. The spontaneous arousal index was 14.0 per hour, and the respiratory arousal index was 0.4 per hour.

**Respiratory disorders**: Respiratory analysis demonstrated 2 obstructive apneas, 0 mixed apneas, and 0 central apneas, with a total of 2 apneas. The apnea index was 0.4 events/h, and apnea-hypopnea index (AHI) was 0.4 events/h. The patient spent 49.0% of total sleep time in the supine position with a supine AHI of 0.8 events/h. The REM AHI was 0.0 events/h versus non-REM AHI of 0.5 events/h.

**Oxyhemoglobin desaturation**: Number of desaturations was 2 with a nadir reaching 96% in stage 2 sleep, followed by apnea, and longest apnea is 26 s occurring in stage 2 sleep.

**Periodic limb movement**: No periodic limb movement was noted.

**Diagnosis**: Patient has no obstructive sleep apnea (AHI = 0.4 events/h).

**Sleep staging summary**

| | | | |
|---|---|---|---|
| Recording start time | 22:50:20 | Recording end time | 05:45:20 |
| Analysis start time | 22:57:20 | Analysis end time | 04:52:19 |
| Total recording time (min) | 415 | Epoch size (sec) | 30 |
| Number of awakenings | 11 | Total sleep time (min) | 322 |
| Wake time after sleep onset (min) | 33 | Sleep efficiency (%) | 77.6 |
| Sleep onset latency (min) | 7 | Stage REM latency (min) | 19 |
| Total sleep period time (min) | 355 | | |

**Respiratory event summary**

| Parameter | REM | Non-REM | Sleep |
|---|---|---|---|
| Apnea | 0 | 2 | 2 |
| Hypopnea | 0 | 0 | 0 |
| Apnea and hypopnea | 0 | 2 | 2 |
| Duration of apnea (min) | 0.00 | 0.85 | 0.87 |
| Duration of hypopnea (min) | 0.00 | 0.00 | 0.00 |
| Duration of apnea and hypopnea (min) | 0.00 | 0.85 | 0.87 |
| Apnea-hypopnea index (events/h) | 0.0 | 0.5 | 0.4 |
| Respiratory arousal index (events/h) | 15.8 | 13.4 | 0.4 |

**Heart rate summary**

| | |
|---|---|
| Average heart rate (beats/min) | 56 |
| Slowest heart rate (beats/min) | 31 |
| Fastest heart rate (beats/min) | 83 |

**Oxyhemoglobin summary**

| Parameter | REM | Non-REM | Sleep |
|---|---|---|---|
| Lowest oxyhemoglobin saturation (%) | 96 | 96 | 96 |
| Mean minimum oxyhemoglobin saturation (%) | 99 | 98 | 98 |
| Mean oxyhemoglobin desaturation (%) | - | - | 2 |

**Arousal summary**

| Type of arousal | REM | Non-REM | Sleep |
|---|---|---|---|
| Spontaneous | 19 | 56 | 75 |
| Respiratory | 0 | 2 | 2 |
| Periodic limb movement | 0 | 0 | 0 |
| Total number of arousals | - | - | 77 |
| Spontaneous (per hour) | 15.8 | 13.4 | 14.0 |
| Respiratory (per hour) | 0.0 | 0.5 | 0.4 |
| Periodic limb movement (per hour) | 0.0 | 0.0 | 0.0 |
| Total number of arousals per hour | - | - | 14.3 |

## Graphical summary of polysomnography (Appendix B)

| Time | 22:50:20 | | | | | | | 5:50:20 |
|---|---|---|---|---|---|---|---|---|
| Hrs | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Epoch | 1 | 121 | 241 | 361 | 481 | 601 | 721 | 841 |

Sleep staging

REM
MOV AWK
1
2
3
4

Respiratory event

Cn.A  +5
Ob.A  +5
Mx.A  +5
Hyp   +5
Uns   +5

Oxyhemoglobin

100
SpO2
40

Body position

F
L
B
R

Periodic limb movement

+10

Arousal