Doctoral Dissertation
博士論文

Study on Classification of Sleeping Breath Sounds and Evaluation of Breathing Quality

(寝息呼吸音の分類と呼吸の質の評価に関する研究)

2023 年 3 月

Wang Lurui
王　魯瑞

Graduate School of Sciences and Technology for Innovation
Yamaguchi University
山口大学　大学院創成科学研究科

Abstract

Sleep is an essential physiological process for the human body. People spend about one-third of their lives sleeping. Both sleep duration and sleep quality are important to human health. Sleep quality describes how restful and restorative the sleep process is. Over 80 sleep disorders are known to affect sleep quality. Among them, sleep-related breathing disorder (SRBD) is the second factor. Sleep-related breathing disorders are sleep disorders in which breathing abnormalities occur during sleep. Abnormal snoring and respiratory arrest or abnormally low breathing during sleep reduce oxygen levels in the blood, increasing the risk of depression, cardiovascular disease, stroke and even death. Therefore, monitoring and analysis of respiration during sleep is gaining increasing importance in healthcare.

Polysomnography (PSG) is considered the gold standard for diagnosing sleep disorders, but PSG is usually performed in an unfamiliar sleep laboratory under the supervision of a medical technician and is often worn with many sensors that interfere with sleep. It is often the case. This research group is developing a breathing sound measurement system that constantly monitors the quality of sleep in a general home environment. This system can easily measure breath sounds during sleep all night with high accuracy without disturbing sleep. The purpose of this research is to develop a technique to classify patterns of breathing sounds and to analyze the quality of breathing in order to more accurately analyze the state of sleep from breath sound information. There are various patterns of sleep breath sounds, such as normal breath sounds and snoring, and abnormal breath sounds and snoring. To develop a method to classify these patterns, to develop an algorithm to calculate ventilation from breath sounds, to estimate the sleep apnea index (AHI), and to assess the quality of breathing during sleep. try.

Specifically, the temporal feature waveform (TCW) is calculated after partly removing the noise of the breathing sounds of sleep with a band-pass filter. Based on the time feature waveform, a respiratory signal effective for analysis is extracted from low-level signals and phase-divided into a respiratory phase and an apnea or low signal. Mel-frequency cepstrum coefficients (MFCC) are then obtained for the respiratory phases, and an agglomerative hierarchical clustering (AHC) algorithm is applied to distinguish between normal/abnormal breathing, normal/abnormal snoring, and normal/abnormal breathing. , tossing and turning, etc., which are less relevant to breathing. The categorized breathing patterns are analyzed every 30 seconds and the relative tidal volume of the breath is calculated. In addition to verifying the effectiveness and accuracy of the technology and analysis method proposed in this study, a method of estimating the apnea syndrome index (AHI) and converting the ventilation volume into high, medium, and low levels, We propose a method to evaluate the quality of breathing in a patient and verify its effectiveness.

This paper consists of six chapters, including an introduction and conclusion.

Chapter 1 introduces the background and overview of this research.

Chapter 2 describes a signal-processing technique for analyzing breath sounds during sleep and a method for classifying breathing patterns. Breathing sound data during sleep often includes disturbed breathing due to bruxism or body movement, ambient environmental noise, etc. In this chapter, the Time Characteristic Waveform (TCW) and the Characteristic Moment Waveform (CMW) are calculated for respiratory sound signals that have undergone preprocessing, such as filtering noise to preprocess the respiratory sounds, and the segmentation of inspiration and expiration is performed. The Mel-Frequency Cepstrum Coefficients (MFCC) are obtained for each respiratory cycle and applied as a feature vector to the Agglomerative Hierarchical Clustering (AHC) algorithm. This method is used to classify ordinary respiratory signals (normal and abnormal breathing, normal and abnormal snoring) from signals less relevant to respiration, such as tossing and turning and environmental noise.

In Chapter 3, using the technology described in Chapter 2, breathing sound data during sleep are classified into apnea, hypopnea, normal breathing, abnormal breathing, normal snoring, and abnormal breathing for each 30-second frame. In addition, we describe a method for classifying events such as no snoring and rolling over and determining the respiratory state.

In Chapter 4, we propose a method for estimating the apnea-hypopnea Apnea-Hypopnea Index (AHI) for classified abnormal breath sounds and low-level breath sound signals, compare it with the diagnostic results of PSG, and examine its validity. And verify usefulness.

Chapter 5 describes a method for estimating ventilation volume from breath sounds. Because normal breath sounds are correlated with ventilation, this study used a quantitative approach to calculate normal breathing and normal snoring and a qualitative method to calculate apnea/hypopnea and abnormal breath sounds. We will propose and compare it with the diagnosis result of PSG and verify its validity.

In Chapter 6, as an application development, an example of applying the breathing sound classification method proposed in this study to heart sound analysis is presented. Finally, we will explain the construction of a data collection distribution system for sharing auscultation data collected at different facilities and hospitals using blockchain technology.

Chapter 7 presents the conclusions and prospects of this study.

要 旨

睡眠は人体にとって不可欠な生理学的プロセスである。人は生涯の約３分の１を睡眠に費やす。睡眠時間と睡眠の質は、どちらも人間の健康にとって重要である。睡眠の質とは、睡眠プロセスがどれだけ安らかで回復力があるかを表すものである。80 以上の睡眠障害が睡眠の質に影響を与えることが知られている。そのうち、睡眠関連の呼吸障害（SRBD）は、２番目の要因となっている。 睡眠関連の呼吸障害は、睡眠中に呼吸の異常が発生する睡眠障害であり、睡眠中の異常ないびきや呼吸停止または異常に低い呼吸などによって血液中の酸素濃度が低下し、うつ病、心血管疾患、脳卒中、さらには死に至るリスクが高まる。したがって、睡眠中の呼吸のモニタリングと分析は、ヘルスケアにおいて益々重要視されている。

睡眠ポリグラフ（PSG）は、睡眠障害診断のゴールデンスタンダードとされているが、PSG は通常、医療技術者の監視の元で慣れない睡眠検査室で行われ、多くのセンサが着けられたため睡眠の妨げとなる場合も多い。本研究グループでは、一般家庭環境下で睡眠の質を常時にモニタリングする寝息呼吸音計測システムを開発している。本システムは、睡眠の邪魔にならないかつ簡便に終夜睡眠の呼吸音を高精度に計測することが可能である。本研究では、呼吸音の情報から睡眠の状態をより正確に分析するため、寝息呼吸音のパターン分類と呼吸の質を解析する技術の開発を目的とする。睡眠呼吸音には、正常な呼吸音といびき、異常な呼吸音といびきなど様々なバターンがある。これらのパターンを分類する方法、ならびに呼吸音から換気量を算出するアルゴリズムを開発することと、睡眠時無呼吸症候群の指数（AHI）を推定すること、さらに睡眠中の呼吸の質を評価することを試みる。

具体的には、寝息呼吸音のノイズをバンドパスフィルターで一部除去したのち、時間特性波形（TCW）を算出する。時間特性波形に基づき、解析に有効な呼吸信号を低レベルの信号から切り出し、呼吸フェーズと無呼吸または低信号にフェーズ分割する。次に、呼吸フェーズに対してメル周波数ケプストラム係数（MFCC）を求め、凝集型階層的クラスタリング（AHC）アルゴリズムを適用して、正常な呼吸/異常な呼吸と、正常ないびき/異常ないびき、また、寝返りなど呼吸に関連性の低いものに分類する。分類された呼吸パターンを 30 秒ごとに分析し、呼吸の換気量相対値を算出する。本研究で提案した技術と解析方法に対して、その有効性と正確性を検証するとともに、無呼吸症候群の指数（AHI）を推定する方法と、換気量を高中低レベルに換算し、睡眠中の呼吸の質を評価する方法を提案し、その有効性を検証する。

本論文は緒論・ 結言を含め ７章から構成されている 。

第 1 章では、本研究の背景と概要を述べる。

第 2 章では、睡眠時の呼吸音を分析するための信号処理技術と呼吸パターンの分類方法について述べる。睡眠時の呼吸音データには、歯ぎしりや体動による呼吸の乱れや周囲の環境ノイズなどが含まれることが多い。本章では、呼吸音を前処理してノイズをフィルタリングなどによる前処理を施した呼吸音信号に対して、時間特徴波形（TCW）と特徴モーメント波形（CMW)を算出し、吸気と呼気の分割を行う。各呼吸サイクルに対して、メル周波数ケプストラム係数（MFCC)を求め、特徴ベクトルとして、凝集型階層的クラスタリング(AHC）アルゴリズムへ適用して、通常の呼吸信号（正常呼吸と異常呼吸、正常いびきと異常いびき）と、寝返りや環境ノイズなど呼吸に関連性の低い信号に分類する信号処理技術を述べる。

第 3 章では、第 2 章で述べた技術を用い、睡眠時の呼吸音データを 30 秒ごとのフレームに対して、無呼吸、低呼吸、正常な呼吸、異常な呼吸、正常ないびき、異常ないびき、および寝返りなどのイベントに分類し、呼吸の状態を判別する方法を述べる。

第 4 章では、分類された異常呼吸音と低レベル呼吸音信号に対して無呼吸低呼吸 Apnea-Hypopnea Index（AHI）の推定方法を提案し、PSG の診断結果と比較検討し、その妥当性と有用性を検証する。

第 5 章では、呼吸音から換気量の推定方法について述べる。正常の呼吸音は換気量に相関性が見られるため、本研究では、正常な呼吸と正常ないびきについては定量的に、無呼吸/低呼吸および異常な呼吸音については定性的に計算する方法を提案し、PSG の診断結果と比較検討し、その妥当性を検証する。

第 6 章では、応用展開として、本件で提案した呼吸音の分類方法を心音解析に応用する事例と、ブロックチェーン技術を用いた異なる施設や病院等で採集した聴診データを共有するためのデータ収集分散システムの構築について説明する。

第 7 章では、この研究の結論と今後の展望について述べる。

contents

Main Technical Term and Notations

Terminology Nouns	Notations	Pages
Sleep-Related Breathing Disorders	SRBD	1
Obstructive Sleep Apnea	OSA	1
Cardiovascular diseases	CVDs	1
Rapid Eye Movement	REM	2
Polysomnography	PSG	2
oxygen saturation	SpO2	2
Respiratory inductance plethysmography	RIP	3
Respiratory Rate	RR	5
Time Characteristic Waveform	TCW	5
Characteristic Moment Waveform	CMW	5
Support Vector Classifier	SVC	6
Apnea-Hypopnea Index	AHI	6
Voice Activity Detection	VAD	6
Mel-frequency cepstral coefficients	MFCC	7
Agglomerative Hierarchical Clustering	AHC	7
Signal-to-Noise Ratio	SNR	13
Principal Component Analysis	PCA	19
Upper Airway Obstruction	UAO	22
Zero-Crossing Rate	ZCR	23
Fast Fourier Transform	FFT	24
Linear Prediction Cepstrum Coefficient	LPCC	26
Discrete Cosine Transform	DCT	26
Automatic Speech Recognition	ASR	26
Non-Rapid Eye Movement	NREM	41
Respiratory Rate Variability	RRV	41
American Academy of Sleep Medicine	AASM	42
mean nocturnal oxygen saturation	MnO2	54
lowest nocturnal oxygen saturation	LoO2	54
Starling Resistor Model	SRM	60
Breath-Sound Amplitude	BSA	60
Blanket Fractal Dimension	BFD	60
Maximum Breathing Pause Interval	MBPI	66
Electrocardiography	ECG	72
Electronic Medical Records	EMRs	79
Practical Byzantine Fault Tolerance	PBFT	81

Chapter 1

Introduction

This introductory chapter introduces the background, review of breathing sound analysis, and the aim of this thesis.

1.1 Research background

Sleep is a necessary physiological process for the human body. People spend approximately one-third of their lifetime on sleep. The amount of sleep needed depends on various factors, such as age, body conditions, or lifestyle. The National Sleep Foundation guidelines advise that the sleep time that healthy adults need is between 7-9 hours[1]. However, enough sleep hours do not necessarily guarantee to get high-quality sleep. Sleep quality and sleep time are both critical for human health. Sleep quality measures how restful and restorative the sleep process proceeds. More than 80 sleep disorders are known to affect sleep quality, divided into seven categories. These categories include insomnia, sleep-related breathing disorders, central disorders of hypersomnolence, circadian rhythm sleep-wake disorders, parasomnias, sleep-related movement disorders, and other sleep disorders[2]. Among all these disorders that cause poor sleep quality, Sleep-Related Breathing Disorders(SRBD) is the second one of all sleep-related disorders(the first one is insomnia)[3].

SRBD is characterized by disordered respiration during sleep, such as abnormal breathing pauses or abnormally low airflow during sleep. SRBD includes many breathing anomalies, including Obstructive Sleep Apnea (OSA), central sleep apnea, sleep-related hypoventilation, and sleep-related hypoxemia disorder. In clinical, a person may be diagnosed with more than one type. OSA is one of the most common and severe sleep-related breathing disorders. OSA is characterized by recurrent complete or near cessation of breathing airflow during sleep. Hypopnea(hypoventilation) is characterized by diminished or restricted respiratory effort with oxygen desaturation. Apnea or hypopnea can last from a few seconds to minutes and occur at least five times or more an hour in adults. Hypoxemia occurs when oxygen desaturation drops and carbon dioxide levels rise. It is often associated with other SRBDs [4].

SRBD is associated with significant medical comorbidities and deteriorates life quality. Many studies have examined the role of SRBD as a pathogenic factor that increases Cardiovascular diseases(CVDs) and cerebrovascular diseases. Blood pressure and heart rate decrease during apnea and an enormous increase at the end of the apnea. These recurrent episodes increase the risk of

cardiovascular diseases and hypertension[5]. SRBD is also accompanied by diabetes, depression, cancer, or sudden death. The common symptoms of OSA are snoring during sleep and excessive daytime sleepiness. SRBD not only cause considerable healthcare cost but also increases the risk of traffic accidents and decreases work efficiency.

SRBD cause less oxygen reaches the brain and blood, negatively affecting the balance of oxygen and carbon dioxide in the blood and causing hypoxia. SRBD has been proven responsible for an exceptionally high heart attack risk and sleep fragmentation. Sleep stages are differentiated by whether or not Rapid Eye Movement(REM) is present. Sleep can be divided into three non-REM stages(N1, N2, and N3) and one REM stage. Stage N3 is also called "deep sleep". Deep sleep is essential for restoring the body and brain, and adequate deep sleep is crucial for body functioning. However, apnea triggers arousal to start awakening to take a breath; these arousals are usually unconscious, but it disrupts the temporal sequence of sleep stages, resulting in significant fragmentation of deep sleep[6]. The fragmentation of deep sleep leads to chronic sleep deprivation and excessive daytime sleepiness. Patients suffering from OSA typically present with depression, cardiovascular disease, stroke, or even death.

Males are more affected by OSA than females, and this prevalence increases with age. Approximately 12% of the adult population in the United States suffer from OSA. In 2015, the cost of diagnosing and treating OSA was about 12.4 billion dollars. 1/7 of the adult population (about one billion) is estimated to suffer from OSA worldwide[7]. Although the diagnosing rate of OSA has increased substantially over the last two decades, it is still highly undiagnosed. According to [7], the estimated cost burden of undiagnosed OSA was more than 149 billion dollars in 2015, mainly including:

- The lost productivity and absenteeism
- Increased risk of comorbidities
- Increased traffic accidents and workplace accidents

Only 12% of respondents reported their concerns about OSA to doctors, and only 30% of patients were warned about the risk of OSA by doctors. About 80% of OSA patients are undiagnosed. Many patients perceive OSA as trivial and do not understand the impact it has on quality of life or refuse the test and treatment because of the expensive cost[8]

The test and diagnosis of SRBD is the first step to treatment. Polysomnography (PSG) is the clinic's golden standard for monitoring and diagnosing OSA. The PSG records physiological signals of the body with various sensors and electrodes, including oxygen saturation (SpO2), brain

waves(electroencephalogram), heart rate(electrocardiogram), leg movements, and breathing airflow[9]. The PSG usually uses the nasal cannula to monitor breathing airflow. The nasal cannula can detect pressure changes in breathing airflow with pressure transducers. Although PSG can record comprehensive body signals and generates an accurate sleeping report, it has several disadvantages in practice. PSG is usually performed in a hospital or sleep laboratory and requires the patient to spend the whole night in the laboratory; the unfamiliar sleep environment may negatively influence sleep quality. The electrodes and sensors attached to the human body may cause uncomfortable during sleep. PSG needs reservation as it needs a medical technician to operate. The PSG usually takes several days, and the costs are expensive. The graphical representation of PSG ias shown in Figure 1.1.

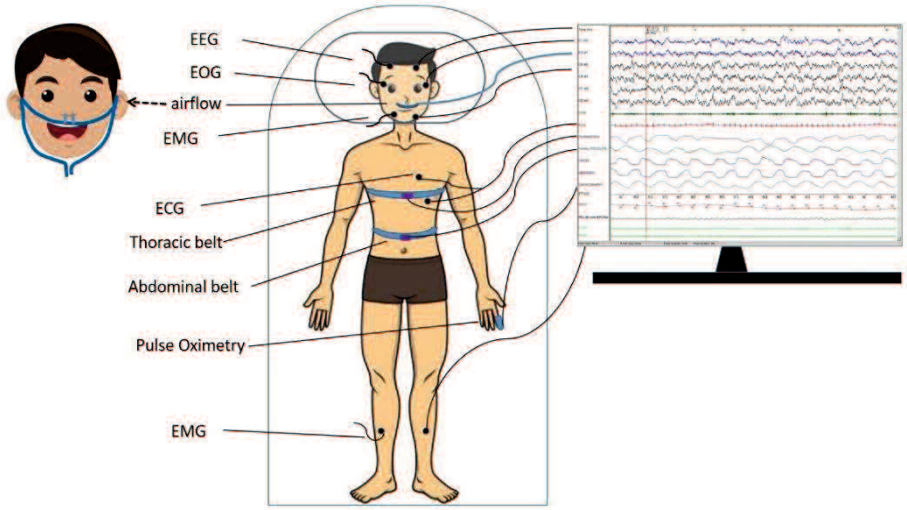

Figure 1.1 The polysomnography and nasal cannual

Researchers also developed other easy-used methods for monitoring breathing during sleep. One frequently used method is the pneumotachograph. A pneumotachograph is usually placed in a mask over the nose and mouth to measure the airflow rate by detecting the pressure change across a small but laminar resistance[10]. Pneumotachograph also causes discomfort, requiring the mask to seal well to prevent air leakage. Another frequently used breath monitoring method is Respiratory inductance plethysmography(RIP). RIP uses two flexible bands attached to the chest and abdomen to measure the chest and abdominal wall movement. It can be used to quantify lung volumes noninvasively[11]. However, the accuracy of RIP in OSA monitoring still needs investigation[12]. Non-contact microphones can also be used to monitor breathing. As SRDB often accompanies snoring, snoring sounds can be used for detecting SRDB early. The microphone is usually placed above or beside a subject's head at an approximate distance of one meter[13]. However, This method is only suitable for people with snoring. The graphical representation of these three methods is shown in Figure 1.2.

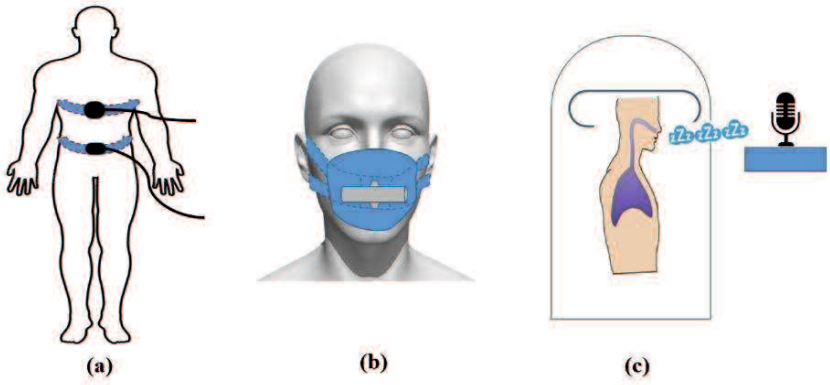

Figure 1.2 (a)RIP (b) pneumotachograph (c) non-contact microphone

With the development of an aging society, the elderly population in Japan is expected to continue to grow. It is estimated that the number of people aged 65 and above will increase to 37.8 billion, or 33% of the total population, in 2060[14]. The need for home medical care is rising, which allows older people to monitor health and diagnose health issues in a home environment rather than a hospital. There is an urgent need to develop new low-cost, easy-to-operate, non-invasive sleep breath monitoring methods that can be used in a home environment. The breath sound-based methods are popular in respiration monitoring as it satisfies the demanding aforementioned and only involve acquiring and processing respiratory sound signals.

1.2 Review of sleep breathing analysis

While breathing, a variety of respiratory sounds are produced. The vibration and turbulence of airflow generate respiratory sounds into and out of the airways and lung tissue. Normal breath sounds can be divided into four specific sounds: tracheal, bronchial, bronchovesicular, and vesicular. Each of the four can be heard over a particular region of the thorax [15]. The auscultation positions of each type of breath sound are shown in Figure 1.3. According to the recording location, respiratory sounds can be classified into lung and tracheal. The lung sounds are detected over the chest wall. The tracheal sounds are detected over the extrathoracic part of the tracheal or at the suprasternal notch. Among respiratory sounds, tracheal sounds are of particular interest in the monitoring of breathing activity and detection of upper airway obstruction because it is easy to detect and can reflect the pathological status of the upper airway.

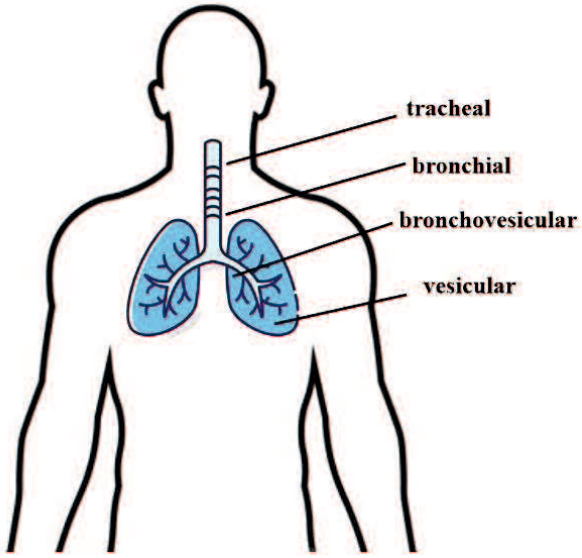

Figure 1.3 Auscultation position of breath sounds

The widely available smartphones and wearable devices also made it possible to monitor respiration during sleep in a home environment. The primary function of these devices is to acquire vital human signs portably in real-time. These devices made monitoring chronic illnesses, such as cardiovascular diseases and SRDB, available in a home environment[16-17]. Monitoring respiratory quality using respiratory sound is becoming a hotspot in recent years. Much research has been done on the acquisition and analysis of respiratory sounds[18-20]. As the Respiratory Rate(RR) is a vital parameter for the human body, some researchers focus on the respiratory rate estimation from breath sounds. Fang et al. proposed a novel sleep respiratory rate detection method based on the Time Characteristic Waveform(TCW) and Characteristic Moment Waveform (CMW) method. A portable and wearable sound device is developed to acquire the breathing sound signal and evaluate the severity of OSA[21]. Nam used a built-in smartphone microphone to record nasal breath sounds and calculate the respiratory rate. The respiratory rate was measured using inductance plethysmography bands and respiratory sounds, respectively. The comparison was performed, and the results show that this method achieved high accuracy even if the smartphone is as far as 30 cm away from the nose[22]. Many researchers developed various wearable devices, either commercially available or at the research state, for detecting and assessing the severity of OSA. Surrel proposed a wearable and energy-efficient system for monitoring OSA during sleep. It is designed as an embedded system using a single-channel electrocardiogram signal and achieved a classification accuracy of up to 88.2% on the PhysioNet Apnea-ECG dataset[23]. Oliver developed a system consisting of non-invasive physiological sensors. Sensors are connected via Bluetooth to a smartphone to store breath sound data. The blood oximeter data are also presented to the user intelligibly[24]. Bsoul developed a real-time sleep apnea monitoring system named Apnea MedAssist, for recognizing OSA episodes with a high degree of accuracy that

can be used in a home environment. Apnea MedAssist uses the patient's single-channel nocturnal ECG to extract feature sets and the Support Vector Classifier (SVC) to detect apnea episodes. It is implemented on smartphones and achieves a classification F-measure of 90% and a sensitivity of 96%[25]. Rosenwein proposed a breath sound analysis method to calculate the Apnea-Hypopnea Index(AHI) as a measure of OSA severity. The breath sounds of 186 adults referred to OSA were recorded with non-contact microphones in laboratory and home environments, respectively. Suspected Apnea/Hypopnea events were automatically detected with the energy envelope of the audio signal. Six features were calculated from supposed periods and classified using a binary random forest classifier; the apnea/hypopnea events can be detected with an accuracy rate of 86.3%[26]. Almazaydeh introduced an automated method to detect sleep apnea based on breath sound signals. The sound signal was filtered and segmented, and a Voice Activity Detection(VAD) algorithm was used to measure the envelope of the signal during breath and breath hold and classify the signal into sound and silence segments[27].

Other researchers focus on the research about the correlation between breath sounds and upper airway narrowing. In [28], bronchial constriction in children was detected by evaluating the sound breath spectrum. Breath sounds acquired near the mouth were analyzed to extract frequency parameters, and the highest frequency of inspiratory breath sounds was calculated. They found that some spectrum curve indexes are not significantly affected by the airflow rate at the mouth, although they successfully indicate airway narrowing. Besides these in-laboratory research, other commercial products also utilize breath sounds to monitor breath quality during sleep[29-31].

1.3 Aim of this dissertation

Although there are many devices or techniques available for breath monitoring during sleep, few of them provide an overall method to monitoring breathing during sleep, especially a few of them focus on the relationship between breathing airflow and sleep quality. As the breath airflow is the direct signal related to the ventilation, it can provide information about oxygen and carbon dioxide balance in the blood. Monitoring breath airflow provides immediate and accurate measuring of breathing quality during sleep. Furthermore, the breath signal acquired in a home environment is more complicated than in a clinical environment. As in the home environment, noise often contaminates the breathing sound signal obtained by sensors. The body movement or sensor friction also causes the signal to be unstable. This research focuses on the analysis of breath sounds for breath airflow estimation and breath quality evaluation during sleep, providing an overall breath quality evaluation and SRBD early detection

method that can be used in a home environment. The primary function of this research is shown in Figure 1.4.

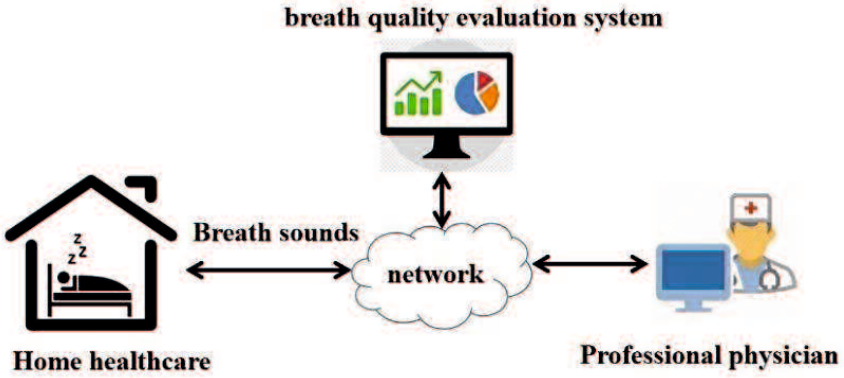

Figure 1.3 The main function of this research

1.4 Outline

Chapter 1 introduces the background and overview of this research.

Chapter 2 describes the sleep-breathing sound classification method. During the breath sound recording, it is often contaminated by noise such as heart sound or ambient noise. First, the breath sound is preprocessed to filter the noise. Second, the breath sound signal is segmented by the Time Characteristic Waveform(TCW) and Characteristic Moment Waveform(CMW). Third, for each breath cycle, the Mel-frequency cepstral coefficients (MFCC) is extracted as the feature vector and classified with Agglomerative Hierarchical Clustering(AHC).

Chapter 3 introduces the sleep-breathing state identification. As the breath sounds are not directly related to breath airflow, it is essential to identify the breathing state based on breath sounds. The whole breathing sound file was segmented into 30-second long clips. Seven breathing states are defined and determined based on the classification result and breath regularity. Each clip is classified into apnea, hypopnea, normal breathing, abnormal breathing, normal snoring, abnormal snoring, and event.

Chapter 4 introduces the Apnea-Hypopnea Index(AHI) estimation method based on breath sound. AHI is calculated with different digital signal processing techniques. The AHI of the whole file is calculated for evaluation of the apnea severity. Subjects with different OSA severity are selected to evaluate the performance of the proposed method.

Chapter 5 introduces the tidal volume estimation method based on breath sounds. Tidal volume is quantitatively calculated for normal breathing and normal snoring states and qualitatively for apnea/hypopnea and abnormal states.

Chapter 6 applies the proposed method to the heart sound analysis. The classification method presented in chapter 2 is used to classify heart sounds into different categories based on the recording quality. A block-chain based data sharing method is proposed for the system development.

Chapter 7 is the summary and conclusion.

References

[1] Hirshkowitz M, Whiton K, Albert S M, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary[J]. Sleep health, 2015, 1(1): 40-43.

[2] Sateia M J. International classification of sleep disorders[J]. Chest, 2014, 146(5): 1387-1394.

[3] Wickwire E M, Collop N A. "Insomnia and sleep-related breathing disorders", *Chest*, Vol.137,no. 6, pp. 1449-1463,2010

[4] Burman D. Sleep disorders: sleep-related breathing disorders[J]. FP essentials, 2017, 460: 11-21.

[5] Penzel T, Kantelhardt J W, Lo C C, et al. Dynamics of heart rate and sleep stages in normals and patients with sleep apnea[J]. Neuropsychopharmacology, 2003, 28(1): S48-S53.

[6] Kimoff R J. Sleep fragmentation in obstructive sleep apnea[J]. Sleep, 1996, 19(suppl_9): S61-S66.

[7] Watson N F. Health care savings: the economic value of diagnostic and therapeutic care for obstructive sleep apnea[J]. Journal of Clinical Sleep Medicine, 2016, 12(8): 1075-1077.

[8] Hidden Health Crisis Costing America Billions, available at https://aasm.org/resources/pdf/sleep-apnea-economic-crisis.pdf. accessed February 05, 2023

[9] Crivello A, Barsocchi P, Girolami M, et al. The meaning of sleep quality: a survey of available technologies[J]. IEEE access, 2019, 7: 167374-167390.

[10] Sun J J, Nanu R, Ray R S. A low cost, simplified, and scaleable pneumotachograph and face mask for neonatal mouse respiratory measurements[J]. Journal of pharmacological and toxicological methods, 2017, 86: 1-11.

[11] Wolf G K, Arnold J H. Noninvasive assessment of lung volume: respiratory inductance plethysmography and electrical impedance tomography[J]. Critical care medicine, 2005, 33(3): S163-S169.

[12] Park D Y, Kim T, Lee J J, et al. Validity analysis of respiratory events of polysomnography using a plethysmography chest and abdominal belt[J]. Sleep and Breathing, 2020, 24: 127-134.

[13] Xie J, Aubert X, Long X, et al. Audio-based snore detection using deep neural networks[J]. Computer Methods and Programs in Biomedicine, 2021, 200: 105917.

[14] Iwata H, Matsushima M, Watanabe T, et al. The need for home care physicians in Japan–2020 to 2060[J]. BMC Health Services Research, 2020, 20(1): 1-11.

[15] Koehler U, Hildebrandt O, Kerzel S, et al. Normal and adventitious breath sounds[J]. Pneumologie (Stuttgart, Germany), 2016, 70(6): 397-404.

[16] Mshali H, Lemlouma T, Magoni D. Adaptive monitoring system for e-health smart homes[J]. Pervasive and Mobile Computing, 2018, 43: 1-19.

[17] Serhani M A, El Menshawy M, Benharref A. SME2EM: Smart mobile end-to-end monitoring architecture for life-long diseases[J]. Computers in Biology and Medicine, 2016, 68: 137-154.

[18] Ionescu C M, Copot D. Monitoring respiratory impedance by wearable sensor device: Protocol and methodology[J]. Biomedical Signal Processing and Control, 2017, 36: 57-62.

[19] Farias F A C, Dagostini C M, Bicca Y A, et al. Remote patient monitoring: a systematic review[J]. Telemedicine and e-Health, 2020, 26(5): 576-583.

[20] Rodday A M, Graham R J, Weidner R A, et al. Predicting health care utilization for children with respiratory insufficiency using parent-proxy ratings of children's health-related quality of life[J]. Journal of Pediatric Health Care, 2017, 31(6): 654-662.

[21] Fang Y, Jiang Z, Wang H. A novel sleep respiratory rate detection method for obstructive sleep apnea based on characteristic moment waveform[J]. Journal of healthcare engineering, 2018, 2018.

[22] Nam Y, Reyes B A, Chon K H. Estimation of respiratory rates using the built-in microphone of a smartphone or headset[J]. IEEE journal of biomedical and health informatics, 2015, 20(6): 1493-1501.

[23] Surrel G, Aminifar A, Rincón F, et al. Online obstructive sleep apnea detection on medical wearable sensors[J]. IEEE transactions on biomedical circuits and systems, 2018, 12(4): 762-773.

[24] N. Oliver and F. Flores-Mangas, “HealthGear: A real-time wearable system for monitoring and analyzing physiological signals,” in Proc. Int. Workshop Wearable Implantable Body Sensor Netw, 2006, pp. 4–64

[25] Bsoul M, Minn H, Tamil L. Apnea MedAssist: real-time sleep apnea monitor using single-lead ECG[J]. IEEE transactions on information technology in biomedicine, 2010, 15(3): 416-427.

[26] Rosenwein T, Dafna E, Tarasiuk A, et al. Breath-by-breath detection of apneic events for OSA severity estimation using non-contact audio recordings[C]//2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015: 7688-7691.

[27] Almazaydeh L, Elleithy K, Faezipour M, et al. Apnea detection based on respiratory signal classification[J]. Procedia Computer Science, 2013, 21: 310-316.

[28] Tabata H, Hirayama M, Enseki M, et al. A novel method for detecting airway narrowing using breath sound spectrum analysis in children[J]. Respiratory Investigation, 2016, 54(1): 20-28.

[29] “Sleeptracker,” 2016. [Online]. Available: http://www.sleepimage.com/

[30] Kelly J M, Strecker R E, Bianchi M T. Recent developments in home sleep-monitoring devices[J]. International Scholarly Research Notices, 2012, 2012.

[31] Doyle D J. Acoustical Respiratory Monitoring in the Time Domain[J]. The Open Anesthesia Journal, 2019, 13(1).s

Chapter 2

Sleep breathing sound Classification

2.1 Introduction to breath sounds

Respiratory sounds are the specific sounds generated by air movement through the respiratory system and are produced by turbulent airflow in the pharynx, glottis, and subglottic region. The sound is transmitted to the body surface through the lung tissue and chest wall. Breath sounds are synonymous with lung sounds, which are essentially the same[1]. Breath sounds are affected by many factors, such as age, gender, breathing pattern, and so on. There are differences in breathing frequency, tidal volume, vital capacity, and gas flow rate among different individuals, and the resulting breath sounds are also different. Breath sounds are directly related to airflow movement, lung shape changes, and airway secretions, which are essential clues for diagnosing lung diseases. It helps diagnose various respiratory disorders and can convey important clinical information about the respiratory system[2]. These symptoms can be diagnosed by hearing the lungs sound with a stethoscope. This diagnosis is called auscultation.

Based on the location of auscultation, breathing sounds can be classified into tracheal, bronchial, bronchovesicular, and vesicular sounds. The tracheal sound heard at the suprasternal notch ranges from 100 Hz to almost 5000 Hz, with a sharp drop in power at a frequency of approximately 800 Hz and little energy beyond 1500 Hz. The lung sound heard over the surface of the chest or back, frequencies range from 100 Hz to 1000 Hz, with a sharp drop at approximately 100 to 200 Hz. Vesicular breath sounds are soft, low-pitched sounds heard over most peripheral lung fields. Bronchovesicular sounds are medium-pitched across the mainstream bronchi, between the scapulae, and below the clavicle. Vesicular sounds are soft sounds heard throughout most lung fields[3]. Table 2.1 provides a summary of normal breath sounds. Abnormal lung sounds include diminished lung sounds and adventitious breath sounds, which various conditions, including asthma, heart failure, or pneumonia, can cause. Pneumonia can cause air or fluid in or around the lungs; emphysema can cause over-inflation of a part of the lungs. Adventitious sounds refer to sounds that are heard in addition to the expected breath sounds mentioned above. The four most adventitious lung sounds are wheezing, stridor, rales, and rhonchi[4]. Normal lung tissue has a similar effect with a low-pass filter as it allows low-frequency sounds to transmit while filtering high-frequency sounds. As the pathological lung tissue is usually occupied by fluid other than air, it can transmit more high-frequency sounds[5-6].

Table 2.1 types and characteristics of normal breathing sounds

Type	Location	Frequency range	pitch
Tracheal	Suprasternal notch	100-5000Hz, energy drop at 800Hz	Loud, high pitch
Bronchial	chest near second and third intercostal space	Similar to tracheal	high
Bronchovesicular	posterior chest between the scapulae and in the center part of the anterior chest	Intermediate between vesicular and bronchial	Intermediate
Vesicular	throughout most of the lung fields	100-1000Hz,energy drop at 200Hz	Low

One normal respiratory cycle can be divided into four primary stages: inspiration, inspiration pause, expiration, and expiration pause[7]. During the inspiration phase, the air flows into the lungs and expands. The inspiratory pause is from the moment that inspiratory airflow stops to the beginning of the expiratory phase. During the expiration phase, the lungs are forced to compress, and the air is forced out. An expiratory pause is from when expiratory airflow stops to the beginning of the next inspiratory phase. A respiratory cycle lasts 4 to 5 seconds for an adult at rest. Inspiration and expiration each require about 2 seconds. A pause of 2 to 3 seconds occurs between each respiratory cycle. The pause can be extended up to 10 seconds[8]. During the pause, there is no air movement into or out of the lungs, and the lung volume is equivalent to the Functional Residual Capacity[9]. A typical respiratory cycle example is shown in Figure 2.1.

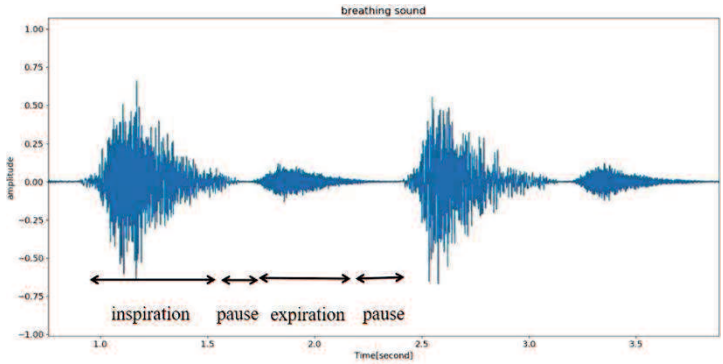

Figure 2.1 typical breathing sound

2.2 breath sound segmentation

Parameters of breath sounds, such as duration, amplitude, and frequency, reflect the physiological conditions of the respiratory system. The loudness and frequency characteristics of breath sounds depend on how narrow the upper airway is and the airflow rate through it. During awake, the neural regulation of the upper airway prevents it from collapsing and keeps it open. However, during sleeping, the muscles that regulate the upper airway relax. This may cause the upper airway to get narrower, and

the velocity of the air moving through it increase, and it has a corresponding effect on breathing sound characteristics and changes the frequency of breath sounds.

Meanwhile, this also causes breath to become quicker and more shallow, especially during the Rapid Eye Movement(REM) stage[10]. The Respiration Rate(RR) or breath rate is the number of breaths a person takes per minute. Normal RR ranges from 12 to 20 times per minute for an adult at rest. A RR under 12 or over 20 breaths per minute while resting is considered abnormal. RR is widely used to monitor the physiology of the human body. It is not only an important indicator of potential respiratory dysfunction but also a key indicator of physiologic processes[11].

The essential step in calculating RR is segmenting the breath sound into breath cycles. Breath sound segmentation is the division and identification of breath sound cycles and the boundaries of each phase. It can provide parameters like the breathing rate and breathing duration and helpful information about the condition of the lung physiology. Most breath sound segmentation algorithms are based on envelope calculation to obtain the changing intensity of breathing. The rule for identifying inspiration and expiration is that the inspiratory pause period is less than the expiratory pause period[12-17]. Research on breath sound segmentation mainly focused on signal preprocessing methods to improve segmentation accuracy under low Signal-to-Noise Ratio(SNR) conditions. Hilbert transform and Shannon entropy are usually used for envelope calculation[18-19]. The common method to segment breath sounds calculates the envelope and sets a threshold for it. Skalicky D proposed a breath phase segmentation method that uses the Hilbert transform to calculate the envelope and find the minimum values of the envelope as the segment points. This method has relatively high robustness and efficiency with low computation[20]. Nam Y proposed a RR estimation method using beath sound recorded by a smartphone. The envelope is calculated by Hilbert Transform, and the envelope is smoothed and bandpass filtered to segment breath sound[21]. Some researchers used Voice Activity Detection(VAD) method or artificial neural network to segment breath sounds[22-23].

However, these methods are only suitable for stable breathing. For people with SRBD, the breathing waveforms are not regular, and the breathing intensity and respiratory rate are highly variable during the whole night sleeping. During recording, ambient noise such as rollover noise or environmental noise may also be recorded. These noises usually have a higher intensity than breath sounds. It is challenging to directly segment the breathing sound based on the original envelope as the breath sound is often covered by high-intensity sound. Recorded breath sounds are preprocessed to remove noise, and the Time Characteristic Waveform(TCW) is used to calculate the envelope[24-25].

2.2.1 breath sound preprocessing

The recording of breath sound signals is easily contaminated by noise, such as white noise, heart sounds, interference caused by friction between sensors and skin, and sudden environmental sounds. These noises inevitably degrade the recording. The first step in analyzing breath sounds is to remove the heart sound interference. Heart sounds are often mixed through the microphone when recording the breath sound. Heart sounds are low-frequency signals, and energy concentrates in [20,200Hz]. It highly overlaps with breath sounds in the frequency range below 200Hz. Many methods have been proposed for heart sound reduction from breath sounds[26-30]. However, most methods are only applied for breath sounds recorded over the chest. As the tracheal sound is used in this research, the interference of heart sound is less to tracheal sound than lung sound recorded on the chest. A bandpass filter is used to remove heart sound. As the primary energy of tracheal sounds concentrates in the frequency range of [50, 2500 Hz], the breath sounds were bandpass filtered by a Butterworth filter with a cutoff frequency of [50,2500Hz]. The filtering process helps reduce the effects of heart sounds and high-frequency noises, meanwhile maintaining the main frequency components of breathing sounds. As the highest filter frequency is 2500Hz, the file was downsampled to 5000Hz according to the Nyquist – Shannon sampling theorem to reduce the file size and speed up the calculation efficiency.

After the heart sound reduction, removing low and high-frequency noises is necessary. Adaptive filtering can eliminate environmental and physiological noise, but it requires multiple sound sources for noise estimation [31]. The white noise and the breathing sound spectrum are overlapped, and orthogonal wavelet decomposition has an adaptive time-frequency localization function, which can be used to distinguish the white noise from the signal mutation part[32]. The Daubechies4 wavelet performs 5-level decomposition, and the visuShrink soft threshold is used. The VisuShrink approach employs a single, universal threshold for all wavelet detail coefficients. A comparison of breath sounds before and after the preprocessing is shown in Figure 2.2.

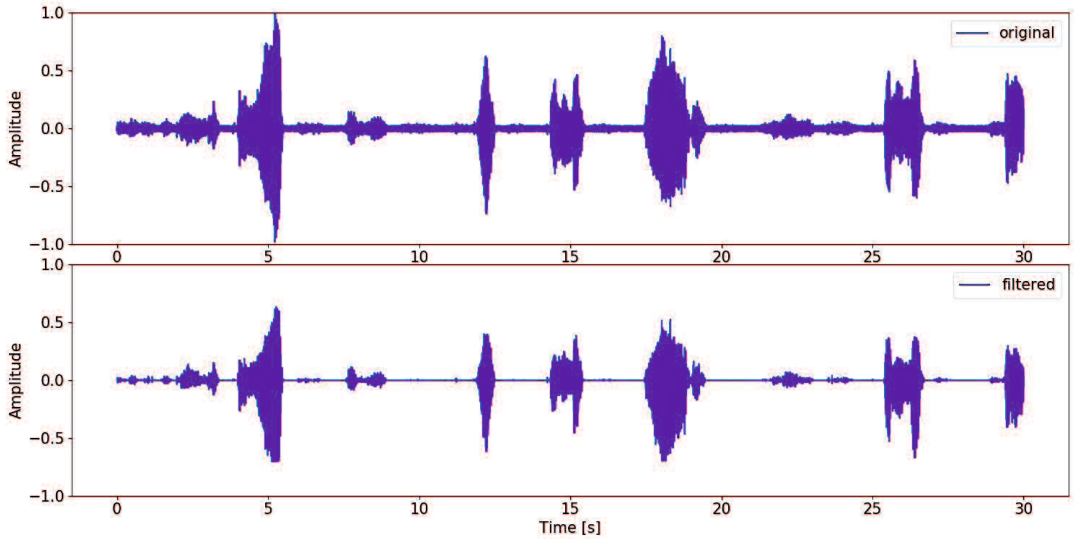

Figure 2.2 breath sound before (top) and after preprocessing(bottom)

2.2.2 breath sound envelope calculation

The calculation of TCW assumes the noise part of the sleep breath sound signal as a signal with zero mean and unit variance. Suppose the sleep breathing sound signal is r(t), the random noise signal is n(t), and the real output signal is y(t) =r(t)+n(t). In a δ neighborhood of time t, the multi-scales characteristic waveform of a signal denoted as c(t, δ), defined as the variance of the output y(t), can be gotten by

$$c(t,\delta) = \int_{t-\delta}^{t+\delta} \left(y(\tau) - \bar{y}(t)\right)^2 d\tau = \int_{t-\delta}^{t+\delta} y(\tau)^2 d\tau - 2\delta\bar{y}(t)^2 \quad (2.1)$$

Where

$$\bar{y}(t) = \frac{1}{2\delta}\int_{t-\delta}^{t+\delta} y(\tau)d\tau$$

The calculation time is proportional to the time scale δ, and the larger δ, the envelope gets smoother. For a signal with length N, the computation of TCW needs 8N additions and multiplications. The normal respiratory duration is 3-5 seconds. So the time scale δ is set to half the breathing duration as 1.5 seconds. A breath sound clip and the corresponding TCW envelope are shown in Figure 2.3.

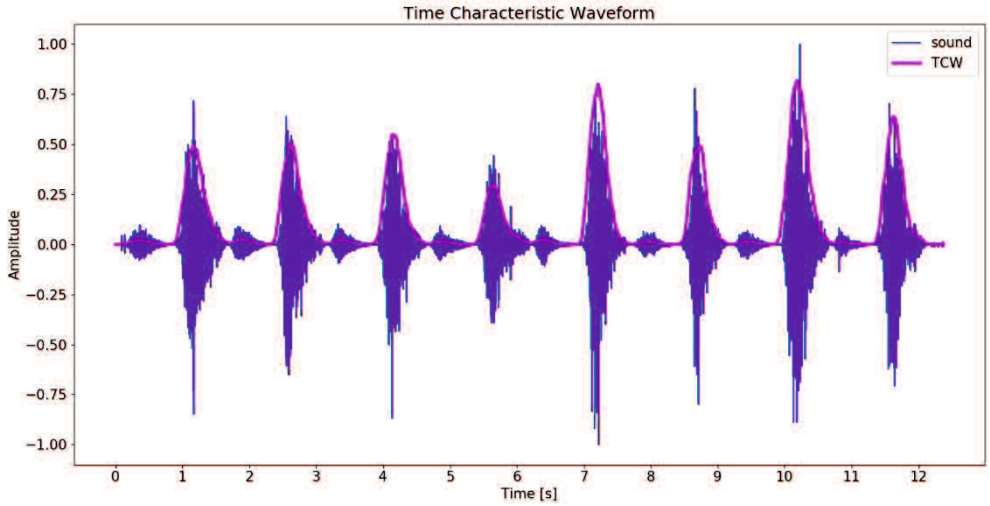

Figure 2.3 TCW envelope of a breath sound clip

Breath sound amplitude is an essential parameter for calculating the envelope. However, snoring or abrupt ambient noise is much louder than normal breathing. High-amplitude snoring sounds often made the normal breathing sounds inaudible. The movement of the body or sensor position also causes the signal amplitude level to change dramatically. In these situations, the normal breathing sound is often misjudged as apnea. Identifying the low amplitude signal and proceeding with special processing is necessary. This research separates the recording signal into normal and low signal parts, and each part is processed using different analysis methods. Apnea is a breathing pause time longer than 10

seconds; therefore, the breathing signal part with a pause time longer than 10 seconds is defined as a low signal part. Otherwise is specified as the normal signal part. For the low-signal part, special processing is needed to extract useful information from a low-intensity signal.

The TCW envelope with a threshold TH_LN is used to separate the breathing sound signal into low signal and normal signal parts. The separate algorithm can be described as following steps:

Step 1:calculate the TCW.

Step 2: Separate the envelope into a low and normal part with threshold TH_LN.

Step 3: merge all normal parts with intervals of less than 10 seconds.

Step 4: merge all low parts with intervals of more than 10 seconds.

The breathing sound signal amplitude is normalized to [-1,1], and the TH_LN is set as 0.05 based on empirical experiments. For the low signal part, it is necessary to extract the breathing sounds with low intensity from the total signal. It is essential to amplify the signal amplitude for the low signal part. The Loudness Normalization extracts the minor signal from the louder signal. Loudness normalization changes the audio intensity to bring the average loudness to a target level[33]. This may cut the peaks that exceed the recording medium's limits. A threshold is needed to set the target loudness level. Based on the empirical experiments, the threshold is set as 0.2. Figure 2.4 shows a signal clip and the signal after loudness normalization. The breath signal from the beginning to 10 seconds has a low amplitude compared with the snoring at 25 seconds; this clip is misjudged as apnea without a normalization process. After Loudness Normalization, the low signal part is enlarged, and the snoring part is cut to the maximum level. The corresponding envelope is amplified for extracting breathing information.

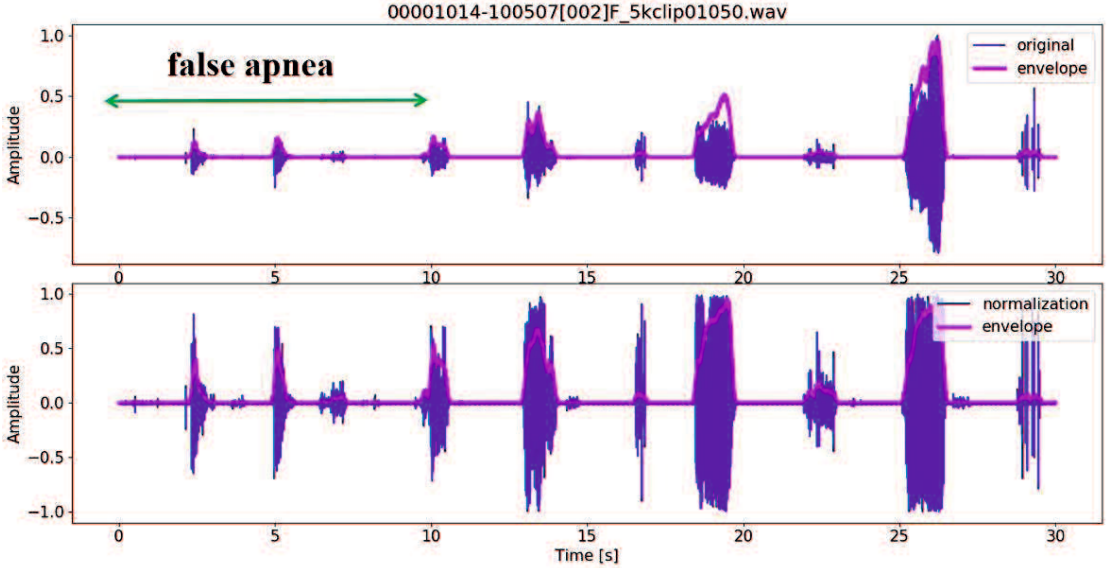

Figure 2.4 low signal fragment before(top) and after(bottom) loudness normalization

2.2.3 breath phase segmentation

The Characteristic moment waveform (CMW) is calculated for breath sound phase segmentation based on TCW. The CMW is derived from the image shape identification in image processing[24]. It is calculated according to Equation 2.2.

$$I(t,\delta,l) = \int_{t-l}^{t+l} (\tau - t)^2 c(\tau,\delta) d\tau \quad (2.2)$$

Where l and is time scale and δ are neighbourhood of time t. The δ is set as half of the breath cycle length. The scale parameters (δ,l) is set as (2.5, 0.1) in this research.

After the TCW and CMW calculation with suitable time scales, breath sound can be segmented as the following steps.

Step 1: Calculate the local maximum point sequence of CMW, mark each point as the breath cycle beginning and end points.

Step 2: Calculate the local maximum point sequence of TCW, mark each point as the center of breath phases.

Step 3: for each local maximum point of TCW, take this point as the center point, search left and right respectively until TCW is less than the TH_LN, mark the stop points as the beginning and end points of inspiration or expiration.

During the tracheal sound recording, the inspiration is usually much louder than the expiration, making expiration hard to detect. Therefore in this research, only the inspiration is segmented and used for the following analysis. The representation of a 30 seconds breath sound segmentation is shown in Figure 2.5. One breath cycle is marked, for example. The breath cycle beginning and end points are marked with a 4-point star, and the center of inspiration is marked with a round point. The beginning and end point of inspiration is marked with a vertical green dotted line. The segment result is shown in Figure 2.6. Each inspiration is marked with the green rectangle.

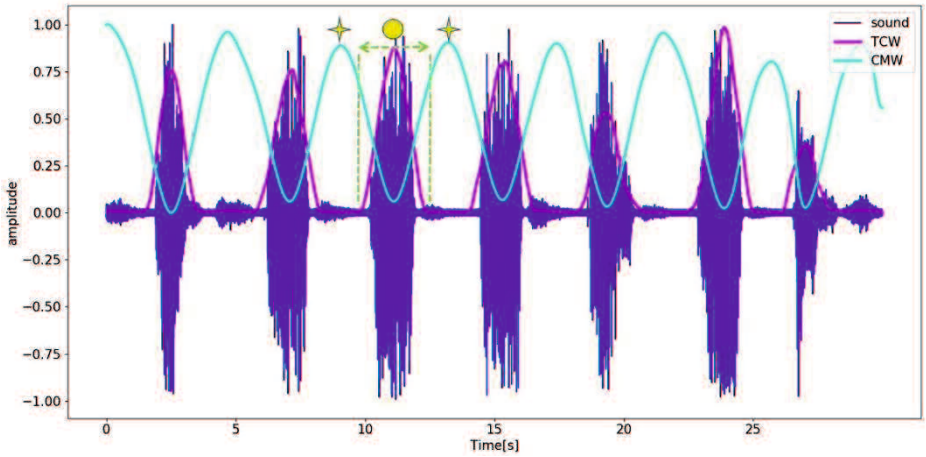

Figure 2.5 the representation of breath sound segmentation

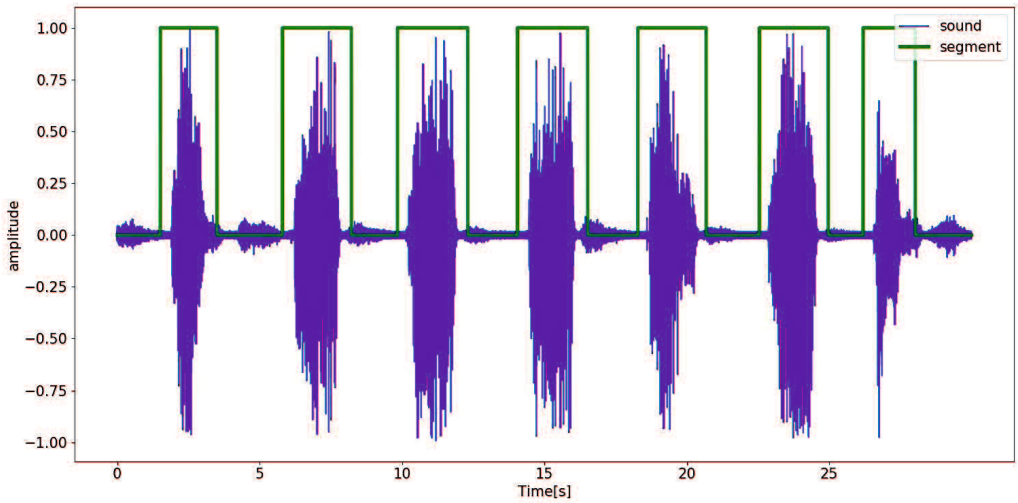

Figure 2.6 breath sound segment result

2.3 breath sound classification

Although the RR can be calculated by breath sound segmentation, the information RR can provide is limited. It does not provide information about the frequency characteristics of each breath cycle. Breath sounds are indicators of lung tissue pathology, the location of secretions in the tracheobronchial tree, and upper airway obstruction—many SRBD symptoms cause differentiation in the breath sound generated during sleep. Previous studies have analyzed the relationship between sleeping breath sounds and the symptom of SRBD. Different symptoms of SRBD have different breath sound patterns or breath sound abnormalities[34]. Snoring is the most common symptom of SRBD. It is the indicator of OSA. In addition to snoring sounds, the obstruction or restricted upper airway often cause irregular breath sounds or disturbed breath rhythms. Labored breathing, gasping, or choking may also be detected during sleep[35]. Recognition and classification of breathing sound patterns and breathing sound abnormalities are important components for lung function diagnosis. In the clinic, this task is usually performed by experts with auscultation. However, for sleep monitoring, the recorded breath sounds usually last several hours, and automatic recognition and classification are needed. Many studies have been done on breath sound recognition and classification. These methods can be classified into supervised methods and unsupervised methods based on whether or not uses the labeled data.

The supervised methods usually consist of two steps: training and testing. During the training step, the labeled datasets is used to train a neural network(or other classifiers) that to classify data or predict outcomes accurately. As the input data is fed into the neural network, it adjusts its weights until it can detect the underlying patterns and relationships between the input data and the output labels, enabling it to yield accurate labeling results when presented with test data. During the test step, a new dataset without labeling is fed to the trained model. The trained model can predict its label[36]. For the breath

sound classification, the spectrograms of the training dataset are often extracted as the input for a neural network. A spectrogram is a visual representation of the frequencies of a signal as it varies with time, able to represent time, frequency, and amplitude all on one graph. The features of the spectrogram, such as the texture and energy distribution, are extracted by the neural network. However, the supervised method needs a large training dataset.

Many studies have been done to classify breathing sounds[37-38]. Bruno proposed a public respiratory sound database for the automatic analysis of respiratory sounds, which includes 920 recordings acquired from 126 participants and two sets of annotations. One set contains 6898 annotated respiratory cycles, some including crackles, wheezes, or a combination of both[39]. Mohammad Fraiwan proposed a dataset that includes sounds from seven ailments and normal breathing sounds. The dataset contains audio recordings from examining the chest wall at various vantage points. These data provide lung sound recordings from 112 Middle Eastern subjects experiencing many pulmonary health conditions[40]. However, these datasets have limited size and are recorded during the daytime. As the breath sounds recorded during sleep, have different characteristics compared with daytime, Whether they are suitable for sleep breath analysis still needs investigation. Some researchers also use deep learning methods to segment breathing sounds[41-45]. However, most of these methods used small-size training datasets that might be insufficient to train and evaluate robust neural networks. Besides the training dataset problem, supervised methods may also encounter the overfitting problem. The representation of the supervised breath sound classification method is shown in Figure 2.7.

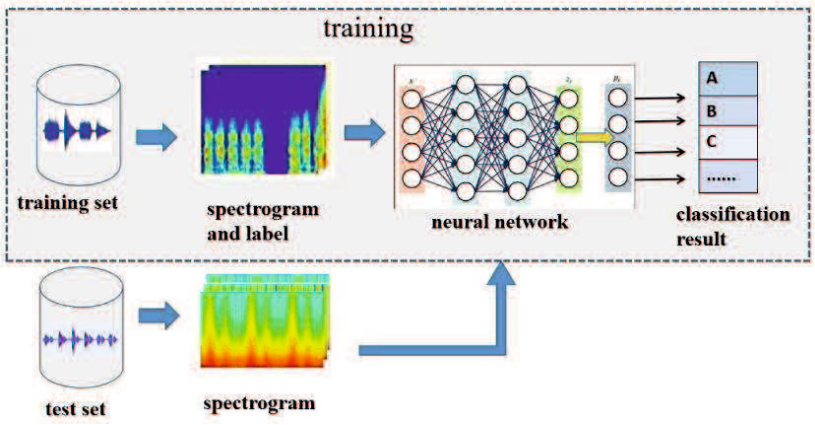

Figure 2.7 The representation of supervised breath sound classification method

The unsupervised method usually segments breath sound into short frames with overlaps and extracts each frame's feature set [46]. The frequently used features include zero crossing rate, spectrogram centroid, spectrogram flatness, etc. The feature set then proceeds with dimension reduction to get low-dimension features. Then it is usually classified with clustering methods. Azarbarzin proposed an automatic and online snore detection algorithm. The Vertical Box algorithm detected the potential snoring episodes, Principal Component Analysis(PCA) was applied to reduce a

10-dimensions into a 2-dimensional feature set, and the K-means clustering algorithm was used to classify all breathing episodes into snore or no-snore categories[47]. Umeki presents an unsupervised respiratory sound classification method based on the subject-adaptation method. The lung sounds are classified into normal/abnormal by a subject-adaptation of acoustic lung-sound models, which achieved an accuracy of 82.7%[48]. The representation of the supervised breath sound classification method is shown in Figure 2.8. These unsupervised methods do not need labeled data sets for training. However, the feature of each frame usually has a low discriminative ability. Breath cycles with different lengths have different frames. This may also have an adverse effect on classification accuracy.

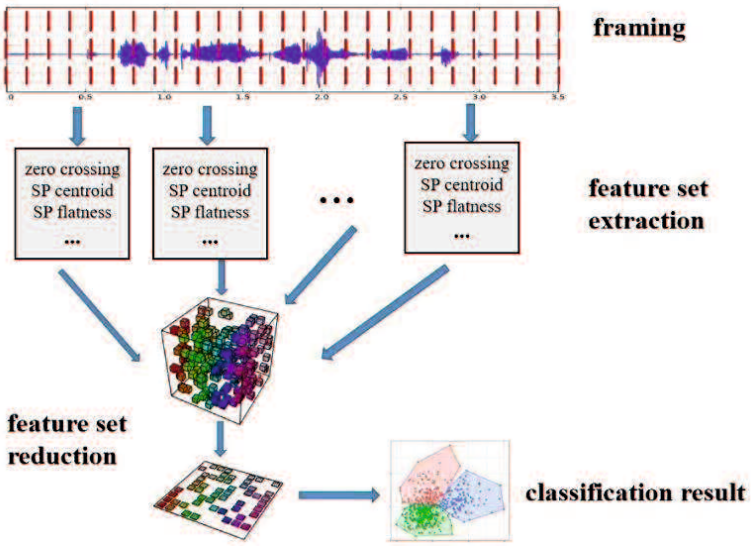

Figure 2.8 The representation of unsupervised breath sound classification method

As traditional supervised and unsupervised breath sound classification methods both have their advantages and disadvantages, a new classification is needed that can utilize both their benefits. Breath sounds vary significantly from person to person. It is challenging to use supervised methods to classify the breath sound, as acquiring a large training dataset is challenging. One night breath sound recording usually lasts several hours and may include thousands of breath cycles. The manual analysis is time-consuming. However, this research focuses on breath ventilation estimation. This requires that the breath cycles be classified into several distinct states, and the breath ventilation can be easily estimated according to these states. Furthermore, breath sounds are non-stationary signals, but a single breathing cycle can be regarded as stationary. As the breath cycle is the basic unit of all-night breathing, it is essential to analyze the characteristics of each breath cycle to identify the breathing states. Based on these assumptions, an unsupervised classification method based on Mel-Frequency Cepstrum Coefficients(MFCC) and Agglomerative Hierarchical Clustering(AHC) is proposed to classify breath cycles into several categories based on their characteristics. MFCC vector is extracted as the feature of each breath cycle. All MFCC vectors form an MFCC matrix as the input for AHC input. MFCC vectors

are clustered into several clusters automatically. It is an unsupervised method with high discriminative ability. The flowchart of the classification method is shown in Figure2.9.

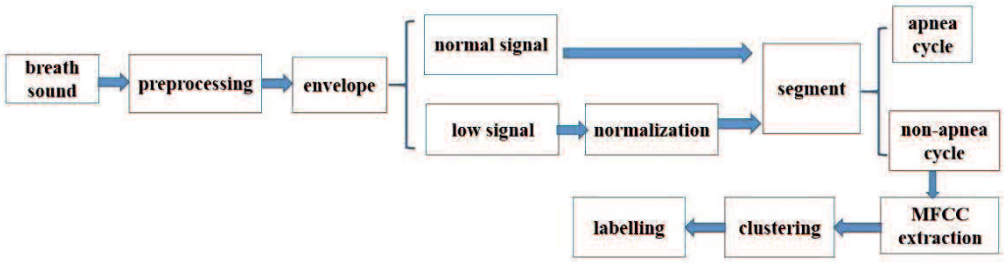

Figure 2.9 The flowchart of the classification method

2.3.1 breath cycle category definition

After the segmentation, the entire breath signal is separated into apnea and non-apnea parts(in section 2.2.2). Non-apnea regions are segmented into breath cycles. For these non-apnea breath cycles, as the breath cycle is the basic unit of sleep breathing signal, it is necessary to identify the characteristics of each breath cycle. These characteristic parameters of each breath cycle include frequency distribution, pitch, breath rate, etc. So that the breathing states, such as normal breathing, snoring, and abnormal snoring, can be identified by these parameters. Breathing patterns may include various states, such as normal breathing, snoring, long and labored breathing, and other irregular breathing rhythms. Breath sounds can be classified into different categories based on various criteria. As this research focuses on breath ventilation estimation, the classification criterion is based on the upper airway states. Several breathing categories are defined based on the spectrum characteristics of breath cycles. These categories are as follows.

(1)Normal breathing

As shown in Figure 2.1, normal breathing sounds have inspiration and expiration stages and clear pauses between them. The frequency range of typical tracheal sound is much broader than normal lung sound, with frequencies ranging between [100,5000]Hz, most of the energy concentrated between [100,1500]Hz, with a sharp drop in energy above a frequency of approximately 800Hz[49]. The frequency variability in inspiration or expiration could be more apparent. The energy distribution in low frequency is smooth, indicating that the respiratory system is in a stable state and the upper airway is not obstructed.

(2)Abnormal breathing

Unlike normal breathing, abnormal breathing sounds usually do not have two stages or a clear pause between inspiration and expiration. A common cause of abnormal breathing is Upper Airway

Obstruction (UAO). It becomes noisy in the case of UAO, demonstrating an increase in the peak spectral power at 1 kHz and an increase in the mean spectral power between [600,1300]Hz[50]. Another common abnormal breathing is breathing with sputum in the respiratory system. Sputum deposition can block the airway and lead to hypoventilation and carbon dioxide retention, which change the spectrum characteristics of breath sounds[51].

(3)Normal snoring

Normal snoring, also called simple snoring, occurs when there is a partial collapse of the throat's soft tissues. Air moves around lax tissue near the back of the throat and causes tissue to vibrate. It often arises from the palate but can also involve other soft tissues of the upper airway. The collapse of normal snoring is incomplete, and the air can flow through the respiratory system to maintain adequate ventilation. As such, simple snoring is generally not considered a health threat[52].

(4)abnormal snoring

Abnormal snoring (also called Apneic Snoring) is caused by partial or complete airway obstruction. Apneic snoring causes a partial or complete airflow stop, resulting in little or no oxygen going to the blood. For the apneic snoring, at the end of the obstruction, the closed upper airway is suddenly opened, and the pressures of the upper and lower airflows are suddenly balanced, causing the upper airway to repeat multiple openings and closings in a short period, producing a popping sound. The collapse degree and resistance of the upper airway may vary significantly from the beginning to the end of inspiration, thus affecting the vibration of the upper airway tissue[53-54]. The snoring sounds in patients with Obstructive Sleep Apnea and simple snoring have different characteristics and effects on breath quality. It is essential to discriminate between these two types of snoring to evaluate the influence on breathing quality.

(4) Uncertain

During the recording, breath sounds are usually contaminated with noises, such as ambient noise or friction noise caused by body movement. When unexpected audio activity is mixed with breathing sounds, the breathing state can not be identified, as the breath signal is usually covered by noise. This type of audio activity is defined as an uncertain category.

2.3.2 breath sound feature extraction

The essential step for classifying breath cycles defined above is feature extraction. Audio features can be classified into time-domain features and frequency-domain features. As this research focuses on

breath ventilation estimation and classification and aims to determine the characteristics of each breath cycle, frequency-domain features are more related to breathing states. Each breathing cycle has different frequency-domain features, such as Zero-Crossing Rate(ZCR), pitch, spectral centroid, and frequency range[55]. However, some categories have similar frequency ranges or spectral centroids, like abnormal and normal breathing. It is necessary to extract features with high discriminative ability. The parts should satisfy the following requirements:

(1) The feature dimensions of each breath cycles with different length should be the same so that the classification algorithm can eliminate the influence of breath duration.

(2) The feature is related to breathing states and has a high discriminative ability. The breathing states defined above can be clearly classified into different categories.

(3) The feature calculation is easy to extract and have low computation cost.

Based on the requirements mentioned above, the Mel-scale Frequency Cepstral Coefficients(MFCC) are used as the feature of each breath cycle. According to research about the human hearing mechanism, the human ear has different hearing sensitivity to sound waves of different frequencies. The human ear has a higher resolution of low-frequency sounds than high-frequency sounds. The Mel scale maps the human auditory perceived frequency to the actual frequency of the sound. By converting the frequencies to the Mel scale, features can better match the human auditory perception[56]. The Mel scale describes the nonlinear characteristics of the human ear frequency, and its relationship with frequency in Hertz can be approximated by Equation 2.3.

$$\mathrm{Mel(f)} = 2595 * \log_{10}\left(\frac{\mathrm{f}}{700} + 1\right) \tag{2.3}$$

f is the frequency in Hertz. The relationship between frequency in Hertz and Mel-scale frequency is shown in Figure 2.10, maximum frequency in Hertz is 20 kHz.

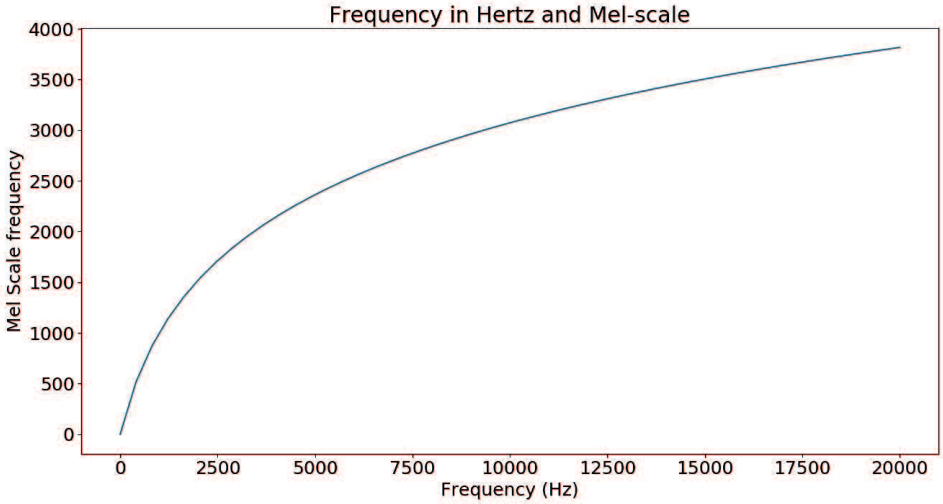

Figure 2.10 the relationship between frequency in Hertz and Mel-scale

The Mel-scale Frequency Cepstral Coefficients(MFCC) is a cepstral parameter extracted in the Mel-scale frequency domain[57]. The MFCC extraction algorithm usually includes the following:

- Pre-emphasis.
- Windowing the signal into frames.
- Applying the Fast Fourier Transform (FFT) on frames to get the short-time Fourier transform spectrum(STFT).

Then the STFT spectrum was filtered with Mel filter banks to get the Mel-spectrum. The Mel-spectrum was transformed into Mel-frequency cepstrum by taking the logarithm and then applying the Discrete Cosine Transform(DCT) to get MFCC coefficients. The MFCC feature describes the power spectral envelope of a single frame. The MFCC extract algorithm is shown in Figure 2.11. The MFCC extraction algorithm is as follows:

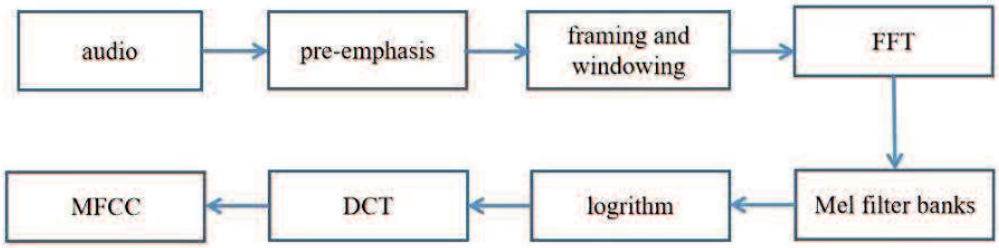

Figure 2.11 MFCC extract algorithm

(1) pre-emphasis

Pre-emphasis processing is passing the signal through a high-pass filter. Applying a pre-emphasis filter to a signal to amplify high frequencies are helpful in several ways. It can balance the frequency spectrum since high frequencies generally have less amplitude than lower frequencies. It can also compensate for the high-frequency part of the voice signal suppressed by the pronunciation system and highlight the high-frequency formant. The pre-emphasis processing can be expressed as the following equation:

$$y(t)=x(t)-\alpha x(t-1) \tag{2.4}$$

x(t) is the original signal, y(t) is the signal after pre-emphasis. α is the filter coefficient, usually set as 0.95 or 0.97 in natural language processing. It is set as 0.97 in this research.

(2) Framing and windowing

Breath signals are short-term stationary signals. After pre-emphasis, it is often necessary to divide the signal into short time frames with overlaps. After the framing, each frame is multiplied by a window function, such as a Hamming window, to increase the continuity of the left and right ends of the frame. A good approximation of the frequency profile of the signal is obtained by concatenating adjacent frames. However, in this research, the breath cycle with different lengths should extract features of the same size. Therefore, the framing step is replaced with breath segmentation in this research. The entire breath cycle signal is taken as one frame to perform FFT.

(3) Power Spectrum

Since it is usually difficult to see the characteristics of the signal in the time domain, it is usually observed by FFT transforming it into an energy distribution in the frequency domain. Different energy distributions can represent different speech characteristics.After FFT, the power spectrum is calculated using the following formula

$$P = \frac{|FFT(x)|^2}{N} \tag{2.5}$$

x is signal of one breath cycle, N is the signal length.

(4) Mel-spectrogram

A Mel-spectrogram is a spectrogram where the frequencies are converted to the mel scale. As mentioned before, the human ear has different hearing sensitivities to sounds of different frequencies. The voice signal between 200Hz and 5kHz greatly influences the hearing system. When two sounds of unequal loudness act on the human ear, the louder sound will affect the perception of the lower loudness sound, making it difficult to detect. This phenomenon is called Audio Masking. Because the lower frequency sound transmits farther on the inner cochlear basilar membrane than the higher frequency sound, the lower frequency sound tends to mask the higher frequency sound. The critical bandwidth for lower-frequency masking is smaller for higher frequencies[58]. Therefore, the researchers designed a set of bandpass filters from low frequency to high frequency according to the size of the critical bandwidth from dense to sparse to filter the input signal[59]. The sound frequency is divided by the critical band, and the voice is divided into a series of frequency groups in the frequency domain to form a filter bank, a Mel filter bank. The Mel filter bank is designed to simulate the non-linear perception of sound by the human ear, with more substantial discrimination at lower frequencies. There are many filters in the low-frequency region, and they are densely distributed, but in the high-frequency region, the number of filters becomes relatively less, and the distribution could be more sparse. The signal energy output by each bandpass filter can be taken as the primary feature of the signal. Since this feature does not depend on the nature of the signal, it does not make any assumptions

and restrictions on the input signal. It utilizes the research results of the auditory model. Therefore, this parameter is more robust than the Linear Prediction Cepstrum Coefficient (LPCC) based on the channel model, more in line with the acoustic characteristics of the human ear, and has better recognition performance when the SNR decreases[60]. The number of filters is usually set between 24-40. In this research, the filter number is set as 32 based on previous research. The 32 Mel filter banks used in this research below 5k Hz are shown in Figure 2.12. Each filter in the filter bank is triangular in shape with a center frequency f(m) where the response is one and decreases linearly towards 0 until the center frequencies of two adjacent filters are reached, and the interval between each f(m) widens as the value of m increases. The frequency is multiplied and accumulated with each filter for the power spectrum obtained after FFT. The obtained value is the energy value of the frame data in the frequency band corresponding to the filter. In this research, 32 energy values were obtained. Since the framing operation is not used in this research, each breath cycle is taken FFT as one frame. Therefore the energy values with the same length are obtained.

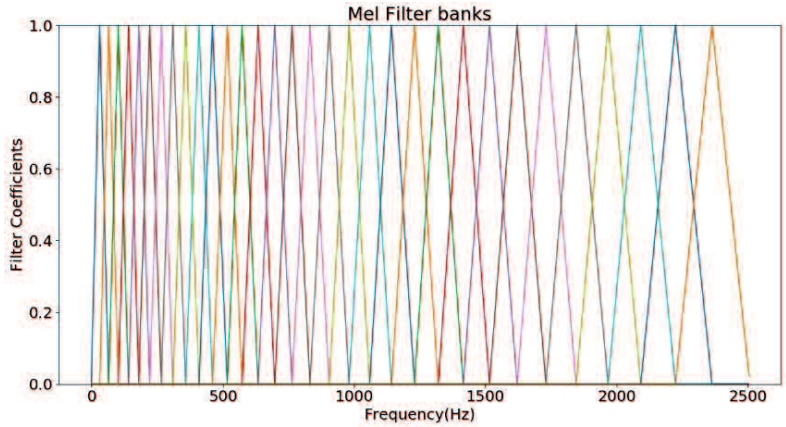

Figure 2.12 32 mel filter banks between 0-2500 Hz

(5) Take logarithm for energies to obtain the log-Mel Spectrogram.

The Mel-spectrogram is usually represented on a log scale. The logarithm is taken on Mel spectrogram to change the unit of energy into decibel.

(6) Discrete Cosine Transform

The log mel-spectrogram computed in the previous step is highly correlated. It can be decorrelated by applying the Discrete Cosine Transform(DCT), resulting in a compressed representation called MFCC. The MFCC of the audio signal is usually a two-dimension matrix, each column present for a frame, and each row in the matrix corresponds to the Mel-frequency cepstral coefficients for the corresponding frame. In this research, each breath cycle is taken as one frame. Therefore a one-dimension MFCC vector is calculated for each breath cycle. For Automatic Speech Recognition (ASR), the resulting cepstral coefficients 2-13 are retained, and the rest are discarded[61].

The reason for discarding the high-order coefficients is that they represent rapid changes in filter bank coefficients, and these fine details do not contribute to ASR.

The first coefficient represents the energy of a frame. It is usually discarded in ASR because it does not carry more information about semantics. However, in this study, the first coefficient is useful in breath sound classification, so the 1-13 cepstral coefficient is kept as an MFCC vector. Each MFCC vector represents the feature of each breath cycle. All the MFCC vectors are concatenated to form a feature matrix. The column is the breath cycle numbers. The row is the MFCC vector length. After transposition, this matrix can be used for classification. The procedure of the MFCC matrix is shown in Figure 2.13.

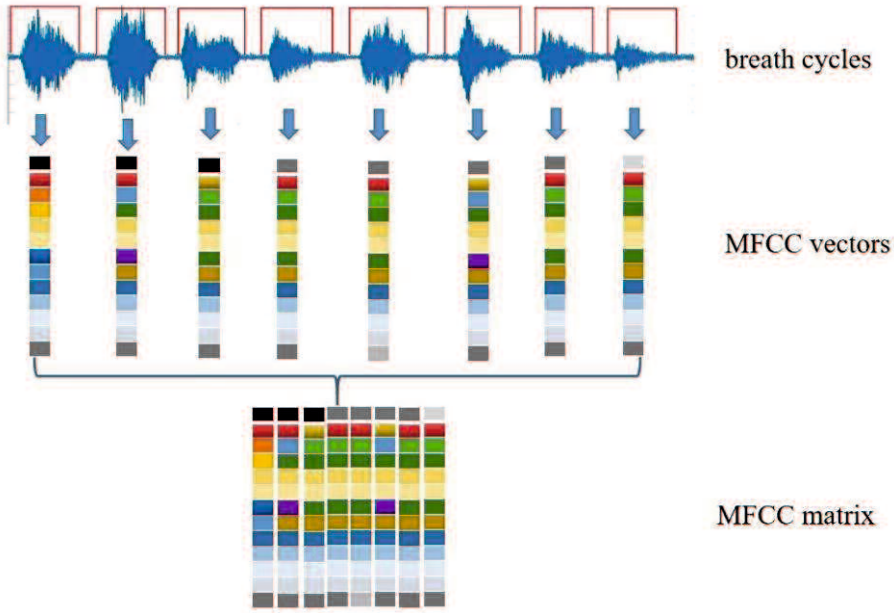

Figure 2.13 feature matrix calculation procedure

2.3.3. Agglomerative Hierarchical Clustering

As a vector can be presented as a point in a high dimension space by its Cartesian coordinates, all the MFCC vectors can be presented as points in a high dimension space. Since similar breath cycles have similar MFCC vectors, the points are also close to each other in space. Therefore, all the points can be separated into several clusters in the space based on the distance between points. The represent of clustering is shown in figure 2.14.

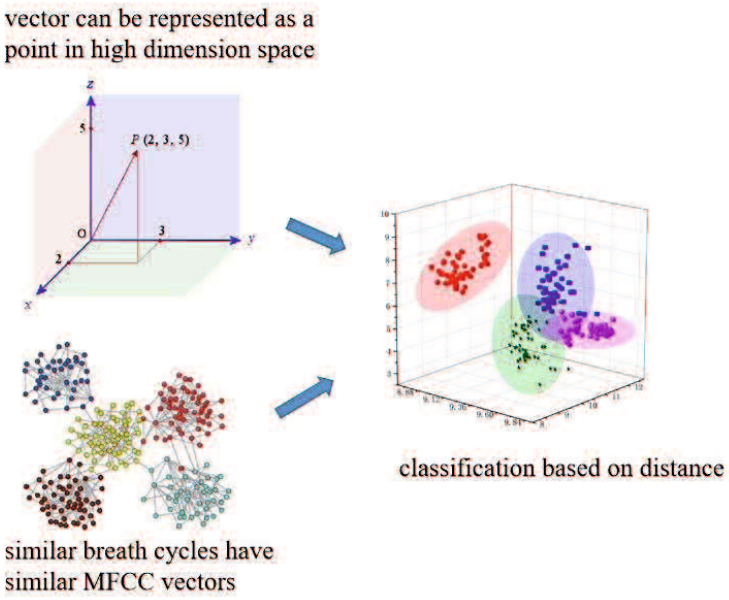

Figure 2.14 MFCC vector clustering based on distance

The distance between these two points can measure as the distance between the two points. Based on our experiences, the Euclidean distance gave the most satisfying cluster result. The Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. In general, if p and q are two points in n-dimensional Euclidean space, then the distance between them can be calculated by Equation 2.5. The calculated Euclidean distance is a number greater than 0. It can be normalized to (0, 1] to better reflect the similarity between users. The similarity of p and q can be calculated by Equation 2.6 and 2.7.

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_1)^2 + \cdots + (p_n - q_n)^2} \tag{2.6}$$

$$\text{Similarity}(p,q)=1/(1+d(p,q)) \tag{2..7}$$

Clustering is an unsupervised machine-learning method that finds natural groups(clusters) of observations based on the internal structure of the data. The input data used for clustering is unlabelled, and the algorithm will discover interesting data structures. Cluster algorithm analysis groups data based solely on the information found in the data that describes the objects and their relationships. The goal is that objects within groups are similar, while objects in different groups are different. The greater the similarity (homogeneity) within groups and the greater the differences between groups, the better the clustering. Hierarchical clustering is a method of cluster algorithm that can discover the structure of the data set in an unsupervised way. The agglomerative method seeks iteratively to merge nodes into bigger clusters. The divisive method splits bigger nodes into smaller ones. Both these two methods can build a hierarchy of all data.

Agglomerative Hierarchical Clustering (AHC) is the most common type of hierarchical clustering[62]. Pairs of clusters are successively merged until all clusters have been merged into one big cluster that contains all objects. Two nodes or clusters with the minimum distance are merged at each iteration. The result is a tree-based representation of all the objects named a dendrogram. The dendrogram may correspond to a meaningful taxonomy. Each horizontal cut of the dendrogram yields a clustering result. AHC needs only a similarity or distance matrix of the dataset for implementation. The procedure of hierarchical clustering is shown in Figure 2.15. The number of clusters needs to be set before the algorithm begins.

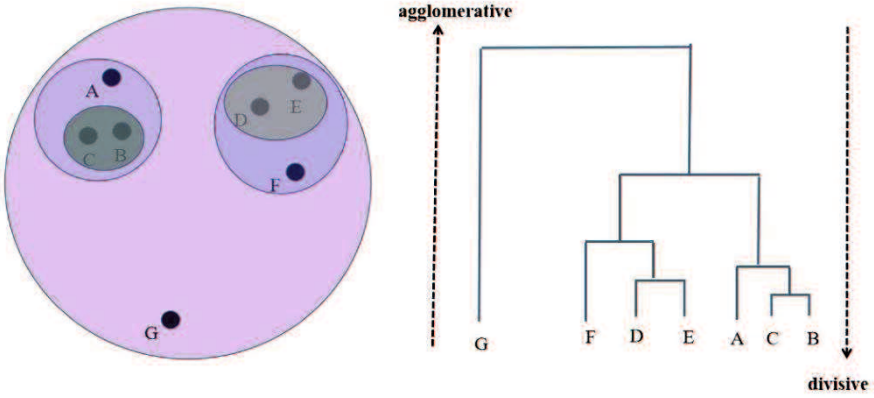

Figure 2.15 Hierarchical clustering procedure

A 120 minutes length breath sound file was selected for demonstration. 592 breath cycles are segmented. The 32 Mel-scale filters are set in MFCC extraction. Based on the structure of the dendrogram. The dendrogram was divided into 7 clusters. The dendrogram is shown in Figure 2.16. The dendrogram is truncated to show the main structure. The breath cycle numbers of each cluster are listed in Table 2.2. The x-axis is the breath cycle index in the file, the y-axis is the Euclidean distance of each MFCC vector.

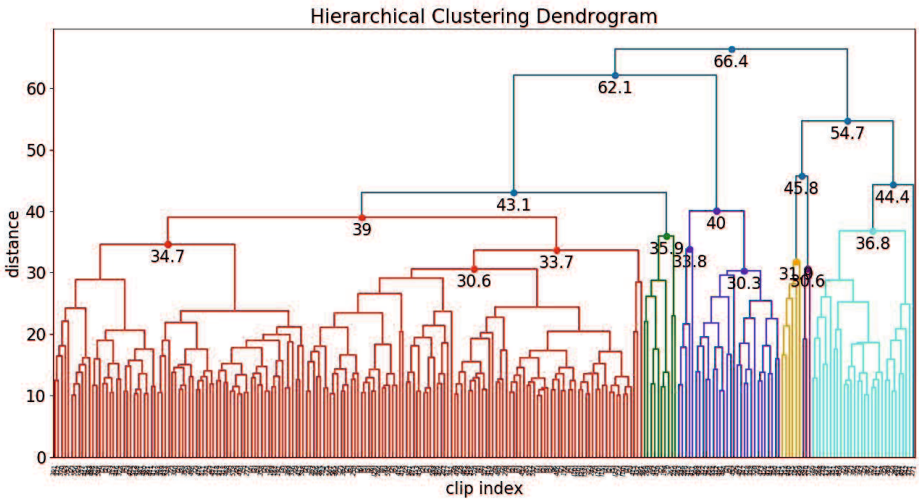

Figure 2.16 dendrogram of cluster result

Table 2.2 property of each cluster

cluster NO.	color	breath cycle number
1	cyan	38
2	black	1
3	Purple	3
4	orange	6
5	blue	34
6	Green	13
7	red	497

2.4. cluster result automatic labelling

The AHC will output a series of labels for each point, representing the clusters it belongs to. Although the clustering algorithm can separate data into clusters, it does not explain each group's nature after grouping. As an automatic model, the working mechanism of clustering is opaque, and the clustering results are challenging to understand and judge[63]. This feature of the clustering algorithm also hinders the subsequent analysis of the clustering results. Therefore, almost all the analysis work on clustering results needs manual processing, which is time-consuming. Especially when faced with high-dimensional data, the time and space overhead of manual analysis of clustering results will increase significantly. Specific to the above breath cycle clustering results, although all breathing cycles are divided into seven categories, the particular characteristics of each category still need manual interpretation.

In recent years, projecting high-dimensional data into a low-dimensional subspace and analyzing it has achieved remarkable results. As the first three dimensions of the MFCC vector have the most prominent ability to describe the envelope of the cepstral spectrum, the first three dimensions of each MFCC vector are used to visualize the MFCC matrix in a 3-dimensional space. The clustering result mentioned above is shown in Figure 2.17. The three axes are the zeroth, first, and second coefficients of the MFCC vector. The color of each point corresponds with the dendrogram.

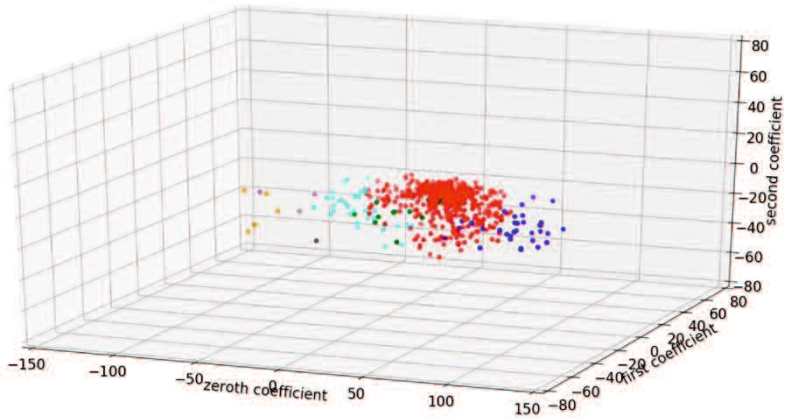

Figure 2.17 cluster result of MFCC matrix in 3-dimensional space

2.4.1. The Zeroth coefficient in clustering

The zeroth coefficient represents the log energy of a frame. It is usually discarded in ASR because it only carries information about the speaker's volume and only carries a little semantic information. However, the log energy of a breath cycle is vital in breath sound classification as it carries information about the upper airway. The snoring breath cycle usually has higher log energy than the breathing cycle. Figure 2.18 shows the points' location from the zeroth coefficient dimension view. Cluster 5, cluster 6, and cluster 7 have larger log energy than clusters 1-4. Therefore, cluster 5, cluster 6, and cluster 7 can be identified as a snoring state, while cluster 1-4 can be identified as a non-snoring state. One snoring cycle with a higher zeroth coefficient is shown in Figure 2.19, and one normal breathing cycle with a lower zeroth coefficient is shown in Figure 2.20.

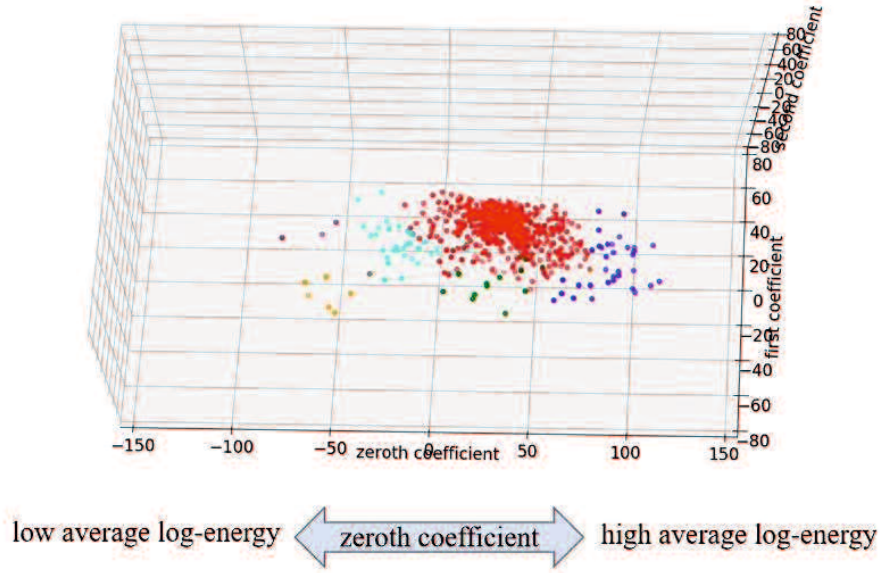

Figure 2.18 cluster result view from zeroth coefficient

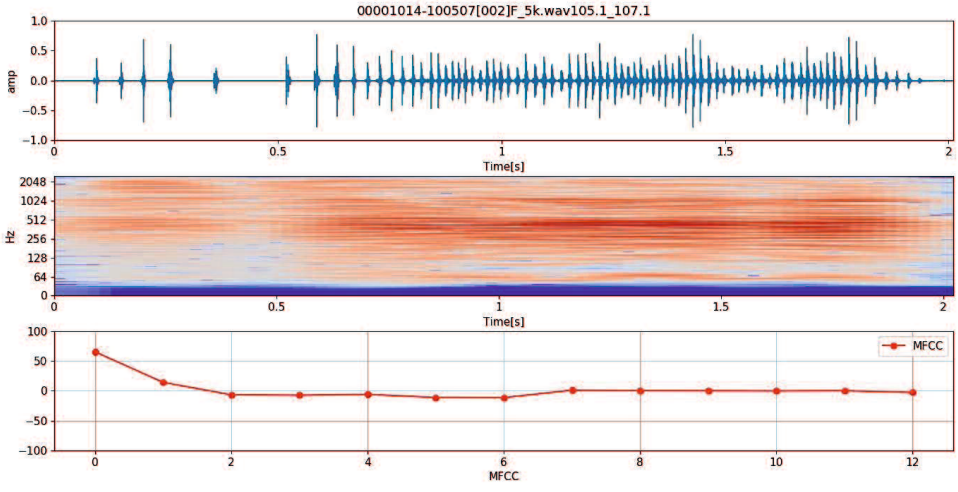

Figure 2.19 snoring cycle with higher zeroth coefficient, (top) waveform, (middle)STFT Spectrum, (bottom)MFCC

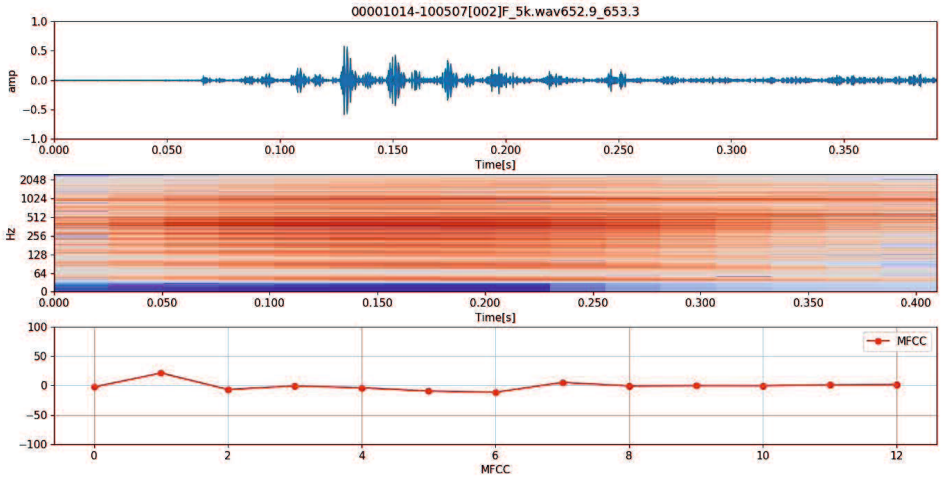

Figure 2.20 breathing cycle with lower zeroth coefficient,(top) waveform, (middle)STFT Spectrum, (bottom)MFCC

2.4.2. The first coefficient in clustering

The cepstral coefficients are a measurement of similarity between the log-Mel cepstral spectrum and the cosine function waves of different frequencies. The first cepstral coefficients capture the primary trend with which the values of this sequence vary. It is also the periodicity of the log-Mel cepstral spectrum envelope. The log-Mel cepstral spectrum of normal breathing has more energy in the low-frequency region and less in the high-frequency region. In other words, the slope of the log-Mel cepstral spectrum has a negative slope. Since the envelope of the log-Mel cepstral spectrum is similar to that of the cosine wave mentioned above, the first cepstral coefficient of normal breathing will have a positive value.

In contrast, the log-Mel cepstral spectrum of abnormal breathing will have a negative value. The reason that abnormal breathing has more high-frequency energy is related to the upper airway state. When the upper airway partially collapses, it gets narrower, and the airflow rate increase to produce more high-frequency sound. The relationship between the log-Mel cepstral spectrum and the first coefficient is shown in Figure 2.21. Figure 2.22 shows the points' location from the first coefficient dimension view. From the first coefficient dimension view, cluster 5 and 6 have more high-frequency energy. Therefore, they can be identified as an abnormal state. While clusters 1-2 and cluster 7 have more low frequency, they can be identified as a normal state.

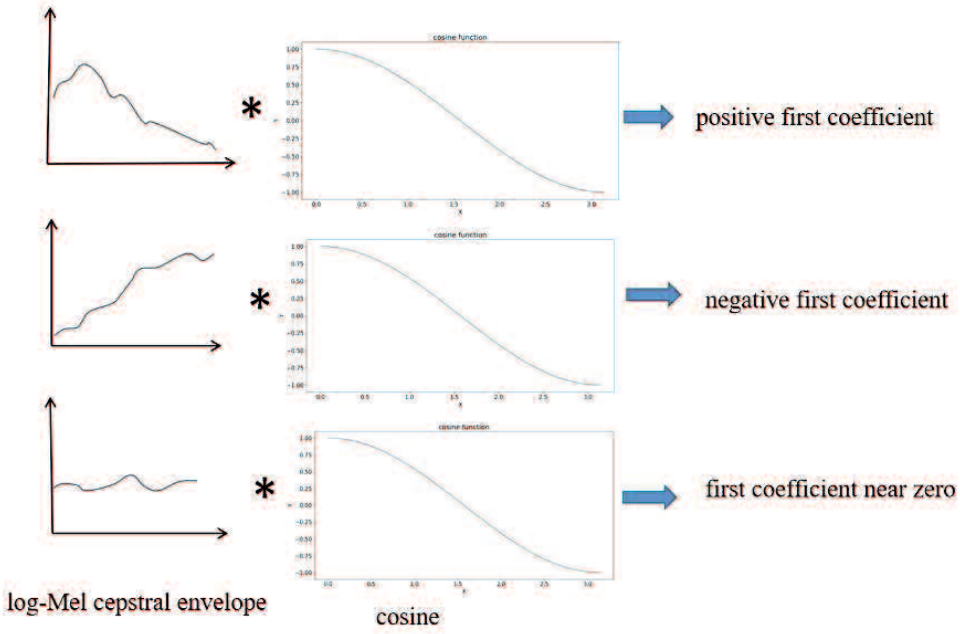

Figure 2.21 the relationship between log-Mel cepstral spectrum and first coefficient

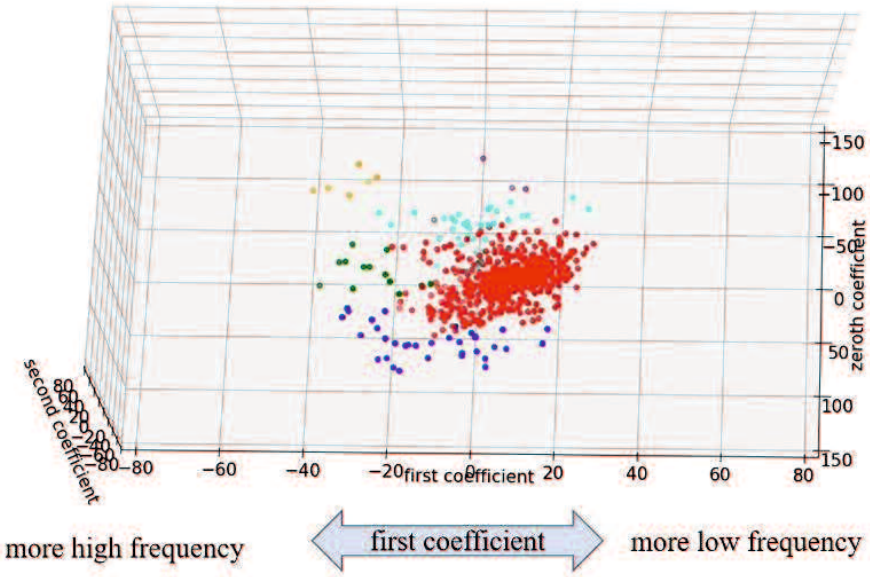

Figure 2.22 cluster result view from first coefficient

2.4.3. The second coefficient in clustering

The second cepstral coefficients capture the proportion of middle frequency and high/low frequency. If the breath sound has more energy in the middle-frequency region and less in the low/high-frequency region, the slope of the log-Mel cepstral spectrum has a negative second coefficient. On the contrary, it has a positive second coefficient. The second coefficient has a smaller ability to distinguish than the first coefficient, and as the coefficients' order increase, the discriminative power gradually decreases. However, the outlier could be easier to identify combined with the second

coefficient. This type of outlier point represents a sound that is not similar to any other clusters. This type of sound usually does not generate by the human respiratory system and can be taken as a noise or uncertain category. Clusters 3 and 4 are outlier clusters away from the space's breathing or snoring clusters. Therefore, all the cluster characteristics can be identified based on location in the space that is composed of the zeroth, first, and second coefficients. The features of each cluster can be summarized in table 2.3. Cluster 3 and cluster 4 can be merged into one cluster that is labeled with uncertain. Cluster 5 and cluster 6 can be merged into one cluster that labeled with abnormal snoring. The relationship between the log-Mel cepstral spectrum and the second coefficient is shown in Figure 2.23. Figure 2.24 shows the points' location from the second coefficient dimension view.

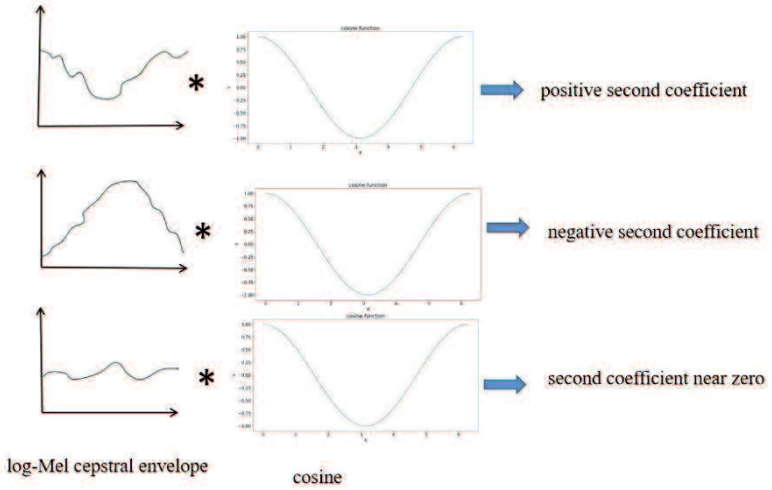

Figure 2.23 the relationship between log-Mel cepstral spectrum and second coefficient

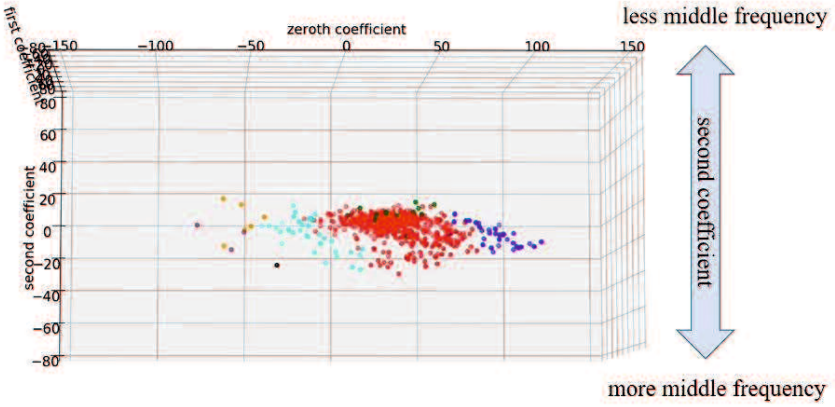

Figure 2.24 cluster result view from second coefficient

Table 2.3 characteristics of each cluster

cluster NO.	color	breath cycle number	category
1	cyan	38	Normal breathing
2	black	1	Abnormal breathing
3	Purple	3	uncertain
4	orange	6	uncertain
5	blue	34	Abnormal snoring
6	Green	13	Abnormal snoring
7	red	497	Normal snoring

A 2 minutes signal clip is selected from the result shown in Figure 2.25 to demonstrate the clustering result. The top is the breath sound signal. The bottom is the clustering results. Each breathing cycle is assigned a label(apnea,normal snoring, abnormal snoring, normal breathing, abnormal breathing, uncertain).

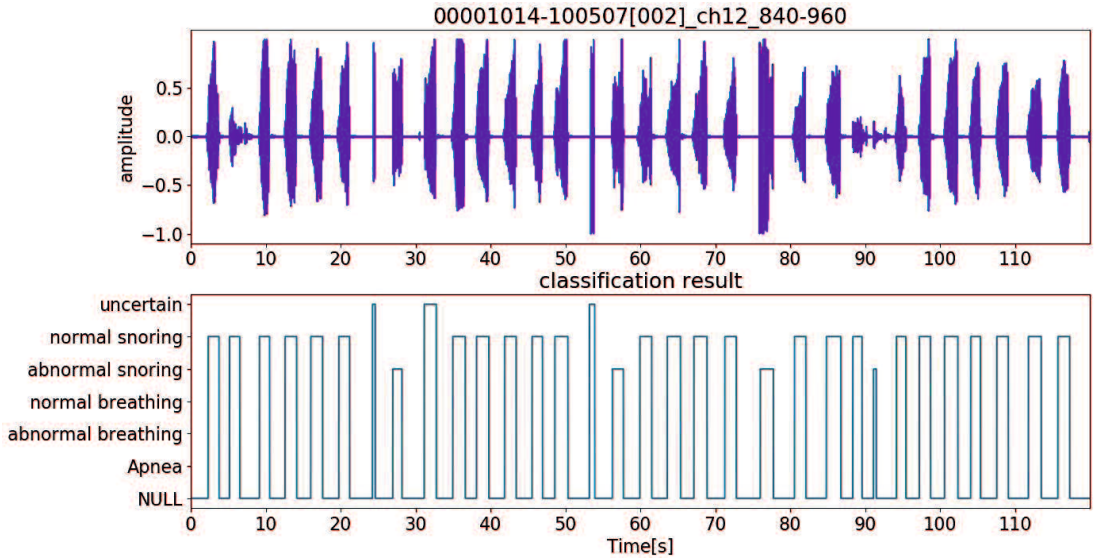

Figure 2.25 cluster result of 2 minute breath signal clip

2.5 Summary

An unsupervised sleep breath sound classification method is proposed to classify breath cycles into normal breathing, abnormal breathing, normal snoring, abnormal snoring, and uncertain categories based on the breathing states. First, the entire signal is separated into low signal and normal signal parts. Next, the low signal is normalized with Loudness Normalization. TCW is calculated as the signal envelope. TCW and CMW are used to segment breath sounds into breath cycles. Apnea is detected in the time domain with TCW envelope. MFCC vectors are extracted as the feature vector for each breath cycle. The MFCC matrix is clustered with AHC. The zeroth, first and second coefficients identify the characteristics of each cluster.

Reference

[1] Pelech A N. The physiology of cardiac auscultation[J]. Pediatric Clinics, 2004, 51(6): 1515-1535.

[2] Sarkar M, Madabhavi I, Niranjan N, et al. Auscultation of the respiratory system[J]. Annals of thoracic medicine, 2015, 10(3): 158.

[3] Bohadana A, Izbicki G, Kraman S S. Fundamentals of lung auscultation[J]. New England Journal of Medicine, 2014, 370(8): 744-751.

[4] Naqvi S Z H, Choudhry M A. An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis[J]. Sensors, 2020, 20(22): 6512.

[5] Zimmerman B, Williams D. Lung Sounds. [Updated 2022 Aug 29]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537253/

[6] Pramono R X A, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: A systematic review[J]. PloS one, 2017, 12(5): e0177926.

[7] Doyle D J. Acoustical Respiratory Monitoring in the Time Domain[J]. The Open Anesthesia Journal, 2019, 13(1)

[8] Hamke E E, Jordan R, Ramon-Martinez M. Breath activity detection algorithm[J]. arXiv preprint arXiv:1602.07767, 2016.

[9] Del Negro C A, Funk G D, Feldman J L. Breathing matters[J]. Nature Reviews Neuroscience, 2018, 19(6): 351-367.

[10] Gould G A, Gugger M, Molloy J, et al. Breathing pattern and eye movement density during REM sleep in humans[J]. Am Rev Respir Dis, 1988, 138(4): 874-877.

[11] Cretikos M A, Bellomo R, Hillman K, et al. Respiratory rate: the neglected vital sign[J]. Medical Journal of Australia, 2008, 188(11): 657-659.

[12] Sierra G, Telfort V, Popov B, et al. Monitoring respiratory rate based on tracheal sounds. First experiences[C]//The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2004, 1: 317-320.

[13] Jin F, Sattar F, Goh D Y T, et al. An enhanced respiratory rate monitoring method for real tracheal sound recordings[C]//2009 17th European Signal Processing Conference. IEEE, 2009: 642-645.

[14] Kulkas A, Huupponen E, Virkkala J, et al. Intelligent methods for identifying respiratory cycle phases from tracheal sound signal during sleep[J]. Computers in Biology and Medicine, 2009, 39(11): 1000-1005.

[15] Jin F, Sattar F, Pwint M. Phase Segmentation of Noisy Respiratory Sound Signals using Genetic Approach[C]//BIOSIGNALS (2). 2008: 122-127.

[16] Azam M A, Shahzadi A, Khalid A, et al. Smartphone based human breath analysis from respiratory sounds[C]//2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018: 445-448.

[17] Castro J, Marti-Puig P. Real-time identification of respiratory movements through a microphone[J]. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 2014, 3(3): 64-75.

[18] Feldman M. Hilbert transform in vibration analysis[J]. Mechanical systems and signal processing, 2011, 25(3): 735-802.

[19] Benitez D, Gaydecki P A, Zaidi A, et al. The use of the Hilbert transform in ECG signal analysis[J]. Computers in biology and medicine, 2001, 31(5): 399-406.

[20] Skalicky D, Koucky V, Hadraba D, et al. Detection of Respiratory Phases in a Breath Sound and Their Subsequent Utilization in a Diagnosis[J]. Applied Sciences, 2021, 11(14): 6535.

[21] Nam Y, Reyes B A, Chon K H. Estimation of respiratory rates using the built-in microphone of a smartphone or headset[J]. IEEE journal of biomedical and health informatics, 2015, 20(6): 1493-1501.

[22] Abushakra A, Faezipour M. Acoustic signal classification of breathing movements to virtually aid breath regulation[J]. IEEE journal of biomedical and health informatics, 2013, 17(2): 493-500.

[23] Hsiao C H, Lin T W, Lin C W, et al. Breathing sound segmentation and detection using transfer learning techniques on an attention-based encoder-decoder architecture[C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020: 754-759.

[24] Yan Z, Jiang Z, Miyamoto A, et al. The moment segmentation analysis of heart sound pattern[J]. Computer methods and programs in biomedicine, 2010, 98(2): 140-150.

[25] Fang Y, Jiang Z, Wang H. A novel sleep respiratory rate detection method for obstructive sleep apnea based on characteristic moment waveform[J]. Journal of healthcare engineering, 2018.

[26] Nersisson R, Noel M M. Heart sound and lung sound separation algorithms: a review[J]. Journal of medical engineering & technology, 2017, 41(1): 13-21.

[27] Gnitecki J, Hossain I, Pasterkamp H, et al. Qualitative and quantitative evaluation of heart sound reduction from lung sound recordings[J]. IEEE transactions on biomedical engineering, 2005, 52(10): 1788-1792.

[28] Canadas-Quesada F J, Ruiz-Reyes N, Carabias-Orti J, et al. A non-negative matrix factorization approach based on spectro-temporal clustering to extract heart sounds[J]. Applied Acoustics, 2017, 125: 7-19.

[29] Falk T H, Chan W Y. Modulation filtering for heart and lung sound separation from breath sound recordings[C]//2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008: 1859-1862.

[30] Molaie M, Moradi M H. Heart sound localization in respiratory sounds based on singular spectrum analysis and frequency features[J]. ETRI Journal, 2015, 37(4): 824-832.

[31] Jatupaiboon N, Pan-Ngum S, Israsena P. Electronic stethoscope prototype with adaptive noise cancellation[C]//2010 Eighth International Conference on ICT and Knowledge Engineering. IEEE, 2010: 32-36.

[32] Sello S, Strambi S, De Michele G, et al. Respiratory sound analysis in healthy and pathological subjects: A wavelet approach[J]. Biomedical Signal Processing and Control, 2008, 3(3): 181-191.

[33] Wolters M, Mundt H, Riedmiller J. Loudness normalization in the age of portable media players[C]//Audio Engineering Society Convention 128. Audio Engineering Society, 2010.

[34] Kim T, Kim J W, Lee K. Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques[J]. Biomedical engineering online, 2018, 17: 1-19.

[35] Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults[J]. Journal of clinical sleep medicine, 2009, 5(3): 263-276.

[36] Hastie T, Tibshirani R, Friedman J, et al. Overview of supervised learning[J]. The elements of statistical learning: Data mining, inference, and prediction, 2009: 9-41.

[37] Reichert S, Gass R, Brandt C, et al. Analysis of respiratory sounds: state of the art[J]. Clinical medicine. Circulatory, respiratory and pulmonary medicine, 2008, 2: CCRPM. S530.

[38] Fernando T, Sridharan S, Denman S, et al. Robust and interpretable temporal convolution network for event detection in lung sound recordings[J]. IEEE Journal of Biomedical and Health Informatics, 2022

[39] Rocha B M, Filos D, Mendes L, et al. An open access database for the evaluation of respiratory sound classification algorithms[J]. Physiological measurement, 2019, 40(3): 03500

[40] Fraiwan M, Fraiwan L, Khassawneh B, et al. A dataset of lung sounds recorded from the chest wall using an electronic stethoscope[J]. Data in Brief, 2021, 35: 106913.

[41] Fernando T, Sridharan S, Denman S, et al. Robust and interpretable temporal convolution network for event detection in lung sound recordings[J]. IEEE Journal of Biomedical and Health Informatics, 2022.

[42] Sun X, Lu Z, Hu W, et al. SymDetector: detecting sound-related respiratory symptoms using smartphones[C]//Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015: 97-10

[43] Palaniappan R, Sundaraj K, Ahamed N U. Machine learning in lung sound analysis: a systematic review[J]. Biocybernetics and Biomedical Engineering, 2013, 33(3): 129-135.

[44] Chen Y, Sun Y, Lv J, et al. End-to-end heart sound segmentation using deep convolutional recurrent network[J]. Complex & Intelligent Systems, 2021, 7(4): 2103-2117.

[45] Acharya J, Basu A. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning[J]. IEEE transactions on biomedical circuits and systems, 2020, 14(3): 535-544.

[46] Xia T, Han J, Mascolo C. Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues[J]. Experimental Biology and Medicine, 2022, 247(22): 2053-2061.

[47] Azarbarzin A, Moussavi Z. Unsupervised classification of respiratory sound signal into snore/no-snore classes[C]//2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, 2010: 3666-3669.

[48] Umeki S, Yamashita M, Matsunaga S. Classification between normal and abnormal lung sounds using unsupervised subject-adaptation[C]//2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). IEEE, 2015: 213-216.

[49] Gavriely N, Palti Y, Alroy G. Spectral characteristics of normal breath sounds[J]. Journal of applied physiology, 1981, 50(2): 307-314.

[50] Yonemaru M, Kikuchi K, Mori M, et al. Detection of tracheal stenosis by frequency analysis of tracheal sounds[J]. Journal of applied physiology, 1993, 75(2): 605-612.

[51] Niu J, Cai M, Shi Y, et al. A novel method for automatic identification of breathing state[J]. Scientific reports, 2019, 9(1): 1-13.

[52] Deary V, Ellis J G, Wilson J A, et al. Simple snoring: not quite so simple after all?[J]. Sleep medicine reviews, 2014, 18(6): 453-462.

[53] Van Brunt D L, Lichstein K L, Noe S L, et al. Intensity pattern of snoring sounds as a predictor for sleep-disordered breathing[J]. Sleep, 1997, 20(12): 1151-1156.

[54] Shen F, Cheng S, Li Z, et al. “Detection of snore from OSAHS patients based on deep learning”, Journal of Healthcare Engineering, 2020

[55] Sharma G, Umapathy K, Krishnan S. Trends in audio signal feature extraction methods[J]. Applied Acoustics, 2020, 158: 107020.

[56] Zhou X, Garcia-Romero D, Duraiswami R, et al. Linear versus mel frequency cepstral coefficients for speaker recognition[C]//2011 IEEE Workshop on Automatic Speech Recognition & Understanding. IEEE, 2011: 559-564.

[57] Han W, Chan C F, Choy C S, et al. An efficient MFCC extraction method in speech recognition[C]//2006 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2006: 4 pp.

[58] Oxenham A J. Mechanisms and mechanics of auditory masking[J]. The Journal of physiology, 2013, 591(Pt 10): 2375.

[59] Kopparapu S K, Laxminarayana M. Choice of Mel filter bank in computing MFCC of a resampled speech[C]//10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010). IEEE, 2010: 121-124.

[60] Misra S, Das T K, Saha P, et al. Comparison of MFCC and LPCC for a fixed phrase speaker verification system, time complexity and failure analysis[C]//2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]. IEEE, 2015: 1-4.

[61] Zheng F, Zhang G, Song Z. Comparison of different implementations of MFCC[J]. Journal of Computer science and Technology, 2001, 16(6): 582-589.

[62] Pellegrini T, Portêlo J, Trancoso I, et al. “Hierarchical clustering experiments for application to audio event detection”, Proceedings of the 13th International Conference on Speech and Computer, 2009

[63] Forina M, Armanino C, Raggio V. Clustering with dendrograms on interpretation variables[J]. Analytica Chimica Acta, 2002, 454(1): 13-19.

Chapter 3

Sleep breathing state identification

3.1 breath sound clipping

During the day, breathing can be controlled, both consciously and unconsciously. As the volitional control and wakeful stimuli wane during sleep, the physiological response differs from that of a wake. The breathing regularity increases with the depth of sleep, and irregular breathing patterns usually only occur during Non-Rapid Eye Movement(NREM)[1]. During sleep, the respiratory function is regulated by a complex interaction between the central neural system and mechanical effectors. Many factors affect the mechanics of breathing and ventilation, further magnify in respiratory disorders. The research on Respiratory Rate Variability (RRV) and respiratory states during sleep are still challenging since it requires sophisticated clinical techniques and continuous monitoring of breathing airflow over many nights[2-4]. These signals are easily affected by body movement, ambient noise, and other types of interference. Therefore, the analysis of breathing sound signals is often hampered by these signal aberrations. Although the method proposed in chapter 2 can classify each breath cycle into different categories, It is still challenging to identify the respiratory states in a certain period by only one single breath cycle. Using several breath cycles in a specific time period other than a single breath cycle is more reasonable for identifying the breathing states. Therefore, several breathing cycles series in a short time are taken as a unit for analysis.

The entire breath sound file is cut into short clips to identify the breathing states in a short period. The duration of the clip length is based on respiratory rate. One clip should be short enough to separate each breathing stage. Therefore the audio signal in one clip is stable. Meanwhile, one clip should contain several breathing cycles so that it is long enough to identify the breathing states. The usual respiratory rate during sleep is 12 to 20 times per minute. One clip is better to contain 5-10 breathing cycles.

Further, considering the apnea time in a severe case, the breathing pause usually lasts more than 20 seconds. The length of 30 seconds to 60 seconds is considerable. Since the PSG signal is continuous, the signal needs to be divided into time segments when analyzing the sleep state through PSG. The hospital technician determines the sleep state by analyzing each signal in the segment. The segments extracted from the continuous PSG signals are called epochs. Technicians judge the sleep stage of each epoch by the ratio of different types of brain waves. The standard epoch length defines 30 seconds in

sleep staging, according to the American Academy of Sleep Medicine (AASM) [5-6]. Therefore, to accord with the AASM definition, in this research, the length of the clip duration is set at 30 seconds.

For evaluation of the performance of the proposed method, the PSG-Audio dataset was used in this research as the data source. The dataset comprises 212 polysomnograms along with synchronized tracheal sound. The dataset contains edf files comprising polysomnogram signals, rml files containing all annotations added by the medical team[7-8]. The edf files contain 20 channels. The data from the nasal cannula pressure(in channel 12) and tracheal sound(in channel 19) are extracted from edf files for analysis. The nasal cannula pressure is a standard method to monitor breathing ventilation during sleep[9-10]. The tracheal sounds are extracted from the PSG-Audio dataset. The corresponding respiratory events(obstructive apnea/mixed apnea/hypopnea) are extracted from rml files. The sampling frequency of nasal cannula pressure and tracheal sound are 100 Hz and 48000 Hz, respectively. A 5 minutes data clip is shown in Figure 3.1. The Y-axis of the top sub-figure is an Arbitrary unit, abbreviations as a.u. The Y-axis of the bottom sub-figure is Hypopnea, Mixed Apnea, Obstructive Apnea, and normal breathing(abbreviation as normal). The basic properties of selected channels in the EDF files are summarized in Table 3.1.

Table 3.1 basic properties of the selected channels in the EDF files of the dataset

Channel ID	Channel Label	Description	Sampling frequency(Hz)	Physical minimum(a.u.)	Physical maximum(a.u.)
12	Flow Patient	Pressure cannula	100	-100	100
19	Tracheal	High-quality contact microphone	48000	-100	100

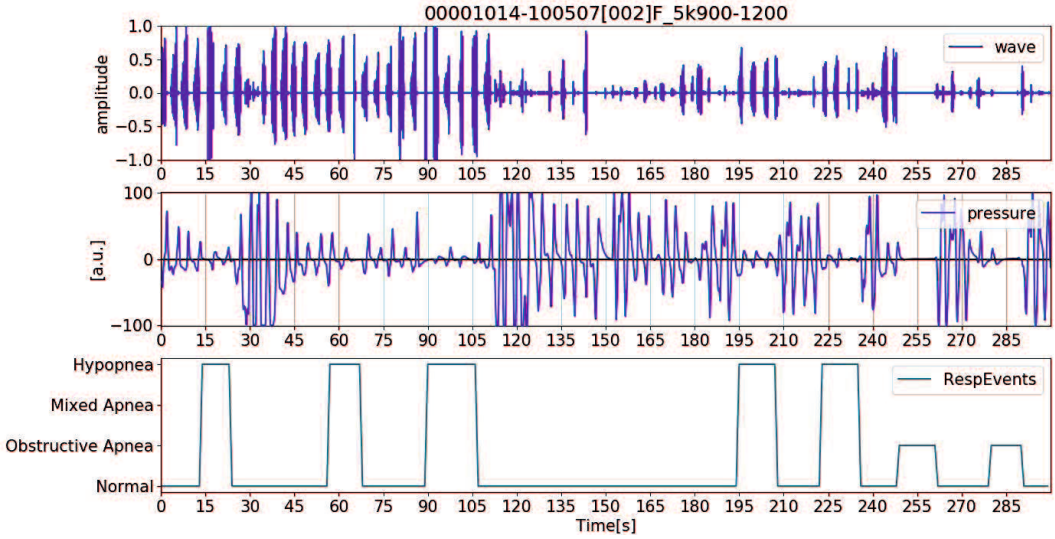

Figure 3.1. data extracted from PSG-Audio (top)tracheal sound, (middle)pressure cannula, (bottom)respiratory events

3.2 breathing states definition and identification

The respiratory pattern refers to breathing rate, depth, and rhythm. Many studies have been done on respiratory pattern analysis. Benchetrit analyzed the diversity of breathing patterns in terms of tidal volume, respiratory duration, and other derived variables and proposed that the variability of breathing patterns could be explained either by a central neural mechanism or by the instability in chemical feedback loops of the respiratory system[11]. Yuan summarized the characteristics of several abnormal breathing patterns, like Kussmaul's breathing and Cheyne-Stokes respiration[12]. However, these studies focused on the respiratory pattern during awake. Hudgel analyzed the breathing pattern variability during sleep with a tightly sealed face mask. They found that the increase in resistance occurred almost entirely above the larynx[13]. However, they only used six healthy adult subjects. As this research focuses on the breathing quality evaluation during sleep, the breath states are defined based on the upper airway states. Similar to the breathing cycle classification method, the breathing states of each clip are defined based on the characteristics of all the breathing cycles in the clip. All the states are defined as follows.

3.2.1. Apnea state identification

According to the definition, breathing pauses for more than 10 seconds during Apnea [14]. Apnea diagnosis is vital in sleep monitoring as it is a common health threat. Apnea deteriorates the balance of oxygen and carbon dioxide in the blood, thus causing a series of comorbidities. After the breathing pause, there is usually followed by several abrupt breathing cycles with apneic snoring. From the aspect of breathing ventilation, the breathing quality of the apnea period is low as the blood oxygen drops dramatically.

Based on the definition, Apnea can be identified in the time domain as the breath cycles have been segmented in chapter 2. The intervals between each inspiration can be calculated based on the segment result. Therefore if the breathing pause time is longer than 10 seconds, the clip is defined as an apnea state. A clip of that identified as Apnea is shown in Figure 3.2.

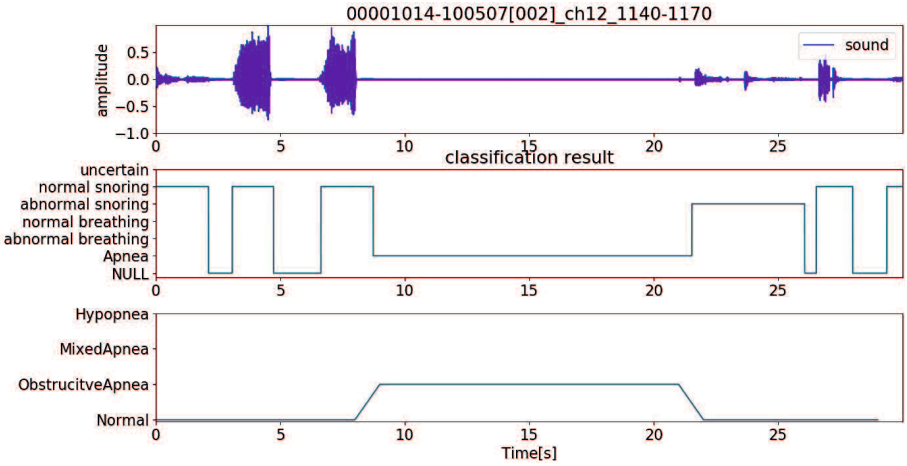

Figure 3.2 a Apnea clip,(top)tracheal sound, (middle)classification result, (bottom)respiratory events label

3.2.2. Hypopnea state identification

Hypopnea is a common symptom of sleep-related breathing disorders. Hypopnea is 10 seconds or more of shallow breathing in which a person's airflow drops by at least 30%. At the same time, blood oxygen levels also drop by at least 3%[15]. According to the American Academy of Sleep Medicine (AASM), the definition of Hypopnea also has a different version[16]. Like the Apnea state, the longer Hypopnea lasts, the more severe its effect on health.

The definition of Hypopnea includes several criteria related to the breathing state. Therefore it is challenging to identify Hypopnea through breathing sounds. As the Hypopnea definition has two criteria: shallow breathing and blood oxygen levels dropping, in this research, two criteria are used to identify the hypopnea state.

(1) Breath interval criterion

As mentioned before, the normal respiratory rate is 12-20 times per minute. Therefore the normal breath cycle duration is 3-5 seconds. As the Apnea definition is respiration stops more than 10 seconds, a shallow breath can be defined as a breath interval more than the normal range but does not exceed the apnea threshold, which is between 5 and 10 seconds. If the shallow breath lasts more than 10 seconds, it can be identified as Hypopnea. A clip of the hypopnea state identified on this criterion is shown in Figure 3.3. The respiration stops between 6-15 seconds. The pause time is between the criterion, thus, is identified as a hypopnea state. This result is in accord with the respiratory events label judged by experts.

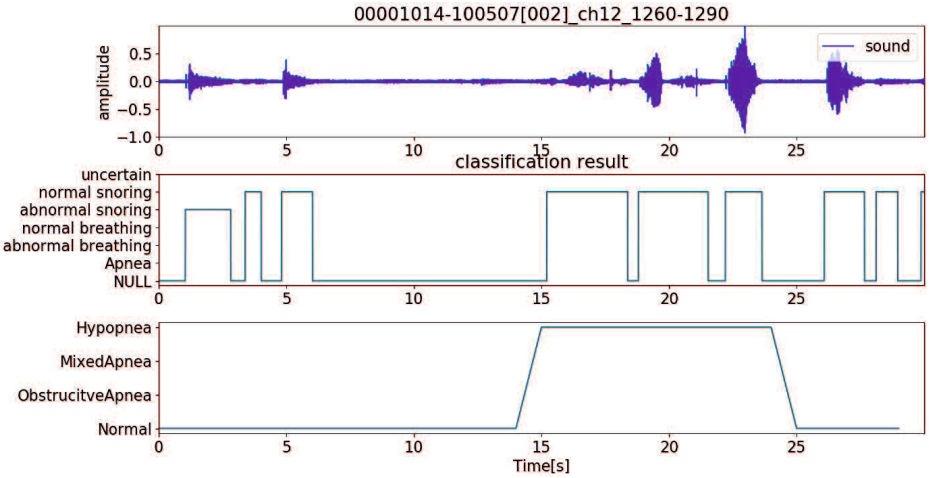

Figure 3.3 hypopnea identified based on breath cycle interval criterion

(2) Abnormal breathing criterion

The other criterion in the hypopnea definition is the blood oxygen levels drop by at least 3%. However, the blood oxygen level does not directly correlate with breath sounds. The blood oxygen level can be estimated from the upper airway state. In contrast, the upper airway state is closely related to the breath sound frequency characteristics. In chapter 2, the breathing cycles related to upper airway partial obstruction are classified as abnormal breathing or abnormal snoring. Therefore, if the abnormal breathing state lasts more than 3/4 proportion, the clip can be identified as a hypopnea state. A clip of the hypopnea state determined based on this criterion is shown in Figure 3.4. In this clip, six breath cycles are segmented. Four breath cycles are classified as abnormal snoring. One breath cycle is classified as abnormal breathing. One breath cycle is classified as normal snoring. Although the breath intervals are in the normal range, the classification result indicates that the upper airway was in an unstable state or partially obstructed. This result is in accord with the respiratory events label judged by experts.

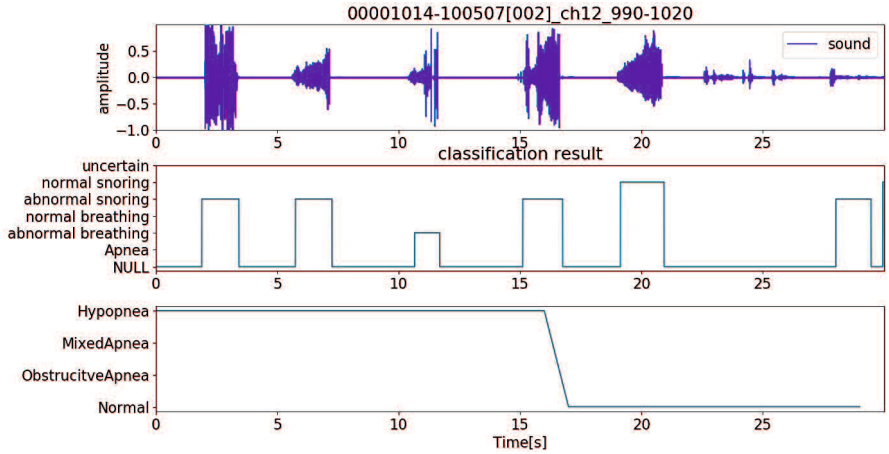

Figure 3.4 hypopnea identified based on abnormal breathing state

3.2.3. Normal breathing state identification

Normal breathing is usually stable breathing during a specific time. During a normal breathing state, the respiratory system and upper airway are stable, and the breathing rate and intensity are stable. However, from the aspect of breathing quality evaluation, the blood oxygen level is not sensitive to slight breathing intensity irregularity[17-18]. Therefore, one or two irregular or abnormal breathing in one clip usually do not affect the breathing quality significantly. Based on this basis, if most of the breathing cycle in a clip is normal breathing based on the classification results, and the breathing rate and intensity are regular, then the clip state is defined as normal breathing. As a clip usually contains 6-10 breathing cycles on average, one or two irregular breathing cycles can not significantly influence the clip's breathing quality. Therefore, the average breathing cycle time proportion threshold is set as 3/4. Clips of typical breathing types are shown in Figure 3.5. The breathing signal in Figure 3.5 is a normal breathing state because all the breathing cycles in this clip are classified as normal, and each breath cycle has regular intensity.

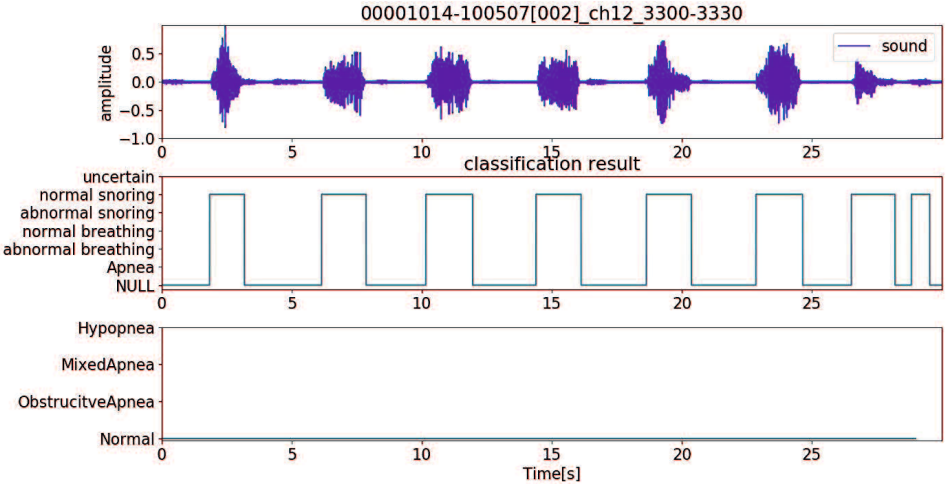

Figure 3.5 normal breathing clip in normal signal

3.2.4. Abnormal breathing state identification

The classification method proposed in chapter 2 can classify breathing cycles as abnormal based on the frequency characteristics. Identifying a clip as abnormal breathing is more challenging as the respiratory rate and intensity must be considered. During an abnormal breathing state, the frequency characteristics of each breathing cycle are different from normal breathing, and the respiratory rate and intensity are also irregular. The definition of abnormal breathing is similar to normal breathing, and the proportion threshold is set between (1/4, 3/4). A clip of the abnormal breathing type is shown in Figure 3.6. Three breathing cycles in this clip are classified as abnormal breathing, as these breathing cycles present different spectrum characteristics compared with regular breathing cycles. Although the respiratory rate is in the normal range, and the intensity and duration of each breathing cycle in this clip

are irregular, indicating that the upper airway is in an unstable state, it does reach the degree that causes hypopnea. Breath states like this can be considered as mild abnormal.

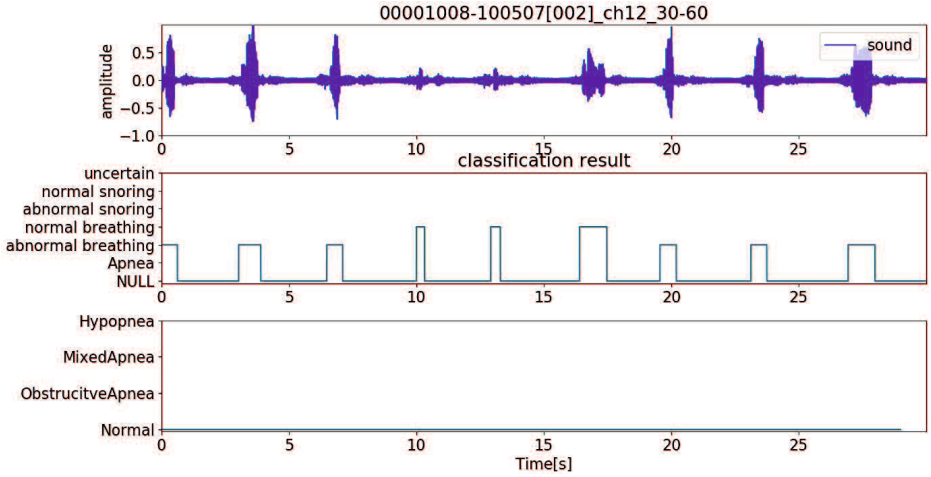

Figure 3.6 abnormal breathing clip

3.2.5. Normal snoring state identification

During normal snoring(simple snoring), the breathing rate and cycle duration are almost the same as normal breathing, only with high intensity. Each snoring cycle presents a clear regular fundamental-harmonic structure. Therefore the definition is similar to normal breathing, and the identification threshold is also the same as normal breathing. A clip of the typical snoring type is shown in Figure 3.7 and Figure 3.8. The breathing signal in Figure 3.7 is a standard normal snoring state. All the breathing cycles in this clip are classified as normal snoring. Each breathing cycle presents a regular intensity and duration. Although there is one unstable breathing cycle in Figure 3.8 based on the classification result, from the aspect of ventilation, this abnormal snoring cycle does not deteriorate the breathing quality significantly. Therefore, the breathing state of this clip can be considered normal snoring.

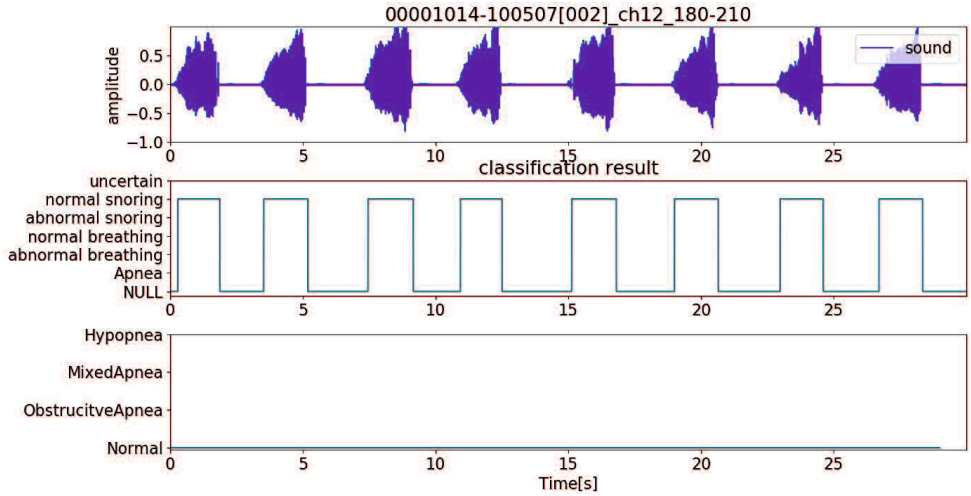

Figure 3.7 normal snoring clip

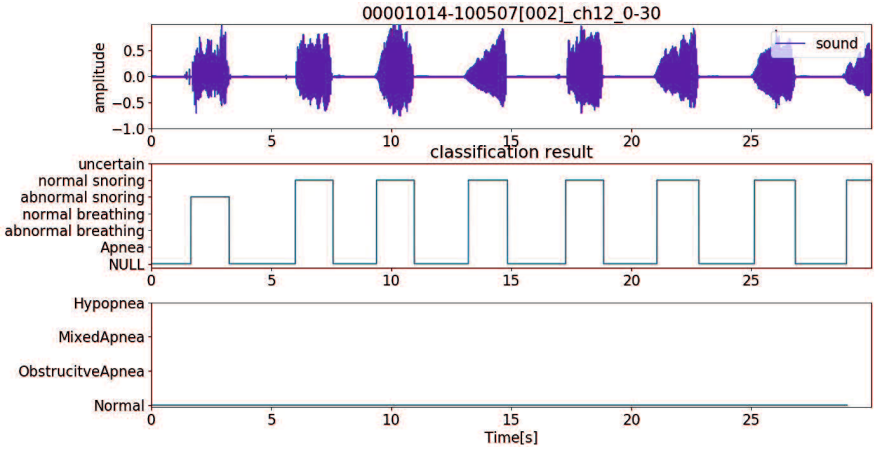

Figure 3.8 normal snoring clip(with one abnormal snoring cycle)

3.2.6. Abnormal snoring state identification

Abnormal snoring is caused by partial or complete obstruction of the upper airway, resulting in an unstable breathing airflow[19-20]. During abnormal snoring, the upper airway is usually obstructive, and the pressures of the upper and lower airflows change dramatically, causing the upper airway to repeat multiple openings and closings in a short period, producing a popping sound. Although the breathing sound intensity is high, the airflow usually drops dramatically. The collapse degree and resistance of the upper airway may vary significantly from the beginning to the end of inspiration, thus affecting the vibration of the upper airway tissue. Abnormal snoring does not present a clear regular fundamental-harmonic structure. The definition is similar to abnormal breathing, and the identification threshold is also set between(1/4, 3/4). A successive abnormal snoring state is often associated with a hypopnea state. A clip of the normal snoring type is shown in Figure 3.9. Based on the classification results, there are four abnormal snoring cycles. These abnormal snoring cycles have varying intensities and spectrum structures. These abnormal snoring cycles do not deteriorate the breathing quality. But the successive abnormal snoring does cause hypopnea.

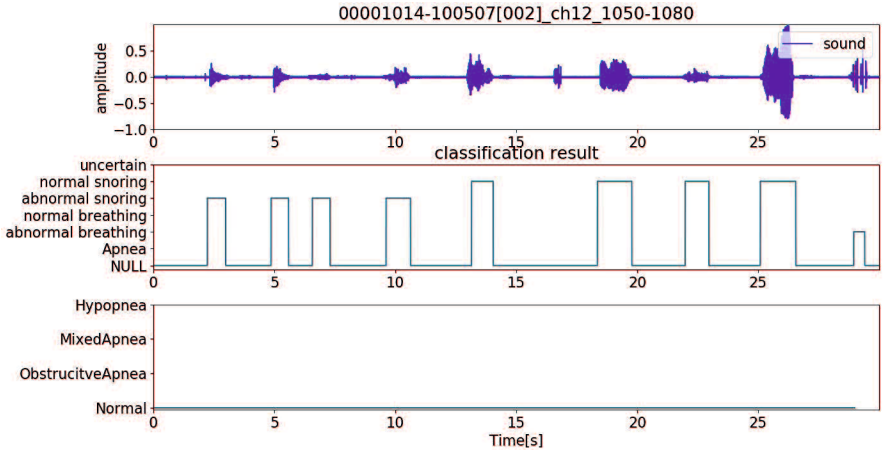

Figure 3.9 abnormal snoring clip

3.2.6. Event state identification

The cluster results of uncertain type are usually ambient noise or unusual events that cause loud sound. The typical cluster results in uncertain categories are turnover noise, sleep talk， cough, bruxism, etc. These parts can be detected and eliminated in speech processing by Voice Activity Detection (VAD)[21-22]. These types of voice activity also can influence sleep quality. Large-scale turning over and stretching of limbs usually occur during a shallow sleep state. However, dedicating the breathing states to these types of activities is challenging. Compared with breathing signals, these activity signals usually have high intensity, and the breathing signal is usually submerged. It is difficult to deduce the breathing states even though a clip has few breathing cycles adjacent to uncertain parts. Clips with long-duration uncertain activity need to be analyzed with special methods that identify the types of activity and the correlation with sleep quality. This research only focuses on the breathing signal to evaluate the breathing quality. Therefore, neglecting these parts in the breathing state identification stages is reasonable. If more than 1/4 of the clip is contaminated by these sounds, the clip is defined as an event. A clip of the ordinary event type is shown in Figure 3.10. The classification of the activity is uncertain as it has a particular spectrum that is unlike any breathing. By checking the sound signal manually, it is most likely the frictional noise caused by turning over. The patient turns over on the bed from 4 seconds to 16 seconds. The recording microphone captures this friction noise. As the existence of this long duration of the noise, the breathing state can not be calculated from this clip.

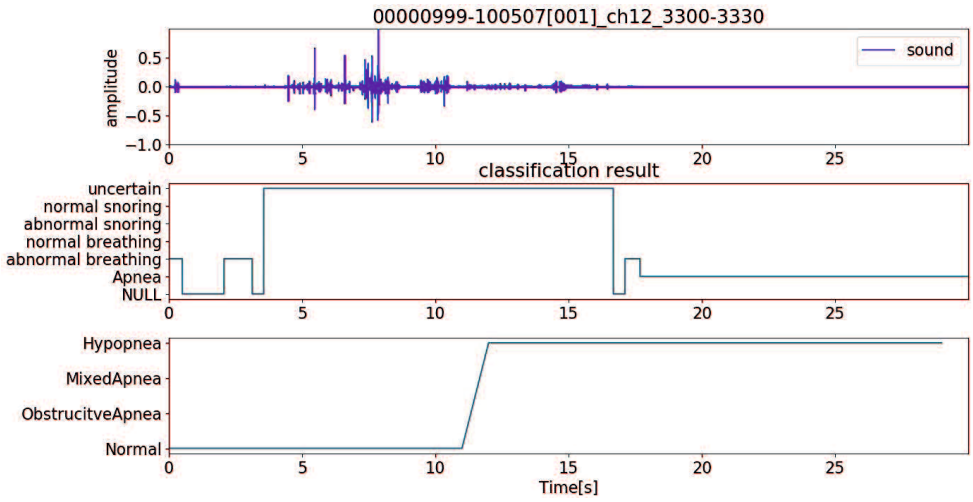

Figure 3.10 event clip

The criterion for clip states identification is summarized in table 3.1. After state identification, all clips are labeled with a state automatically. Therefore, the whole night breathing states variation can be obtained. This result is used for the sleep quality evaluation in chapter 4 and chapter 5.

Table 3.1 the criterion of clip states identification

NO.	state	criterion
1	apnea	Breathing pause time>10 second
2	hypopnea	7<breathing pause time<10 , or abnormal breathing lasts more than 10 seconds
3	abnormal Snoring	abnormal Snoring cycles time proportion between (1/4, 3/4)
4	normal Snoring	normal Snoring cycles time proportion>3/4
5	abnormal Breathing	abnormal Breathing cycles time proportion between (1/4, 3/4)
6	normal Breathing	normal Breathing cycles time proportion>3/4
7	event	Uncertain cycles time proportion>1/4

3.3 Summary

To evaluate the breathing quality in a short period, the breath sound recording file was cut into a 30-seconds length clip. The classification results and the regularity of each breathing cycle identify the clip state. Apnea is identified by a breathing pause time of more than 10 seconds. Hypopnea is identified by a breathing pause time of more than 7 seconds and less than 10 seconds, or abnormal breathing lasts more than 10 seconds. Normal breathing, abnormal breathing, normal snoring, and abnormal snoring are identified by a cycle time proportion threshold of 3/4. The event is identified by an uncertain activity time proportion of 1/4. Each clip is labeled with a state. The whole night breathing states variation can be obtained by concatenating all the state labels.

Reference

[1] Newton K, Malik V, Lee-Chiong T. Sleep and breathing[J]. Clinics in Chest Medicine, 2014, 35(3): 451-456.

[2] Gutierrez G, Williams J, Alrehaili G A, et al. Respiratory rate variability in sleeping adults without obstructive sleep apnea[J]. Physiological reports, 2016, 4(17): e12949.

[3] Vielle B, Chauvet G. Delay equation analysis of human respiratory stability[J]. Mathematical biosciences, 1998, 152(2): 105-122.

[4] Younes M, Ostrowski M, Atkar R, et al. Mechanisms of breathing instability in patients with obstructive sleep apnea[J]. Journal of Applied Physiology, 2007, 103(6): 1929-1941.

[5] Consens F B, Chervin R D, Koeppe R A, et al. Validation of a polysomnographic score for REM sleep behavior disorder[J]. Sleep, 2005, 28(8): 993-997.

[6] Moser D, Anderer P, Gruber G, et al. Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters[J]. Sleep, 2009, 32(2): 139-149.

[7] Korompili G, Amfilochiou A, Kokkalas L, et al. “PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies”, Scientific data, Vol.8, no.1, pp.1-13,2021

[8] Korompili G, Kokkalas L, Mitilineos S A, et al. Detecting Apnea/Hypopnea Events Time Location from Sound Recordings for Patients with Severe or Moderate Sleep Apnea Syndrome[J]. Applied Sciences, 2021, 11(15): 6888.

[9] Nielsen K R, Ellington L E, Gray A J, et al. Effect of high-flow nasal cannula on expiratory pressure and ventilation in infant, pediatric, and adult models[J]. Respiratory care, 2018, 63(2): 147-157.

[10] Thurnheer R, Xie X, Bloch K E. Accuracy of nasal cannula pressure recordings for assessment of ventilation during sleep[J]. American journal of respiratory and critical care medicine, 2001, 164(10): 1914-1919.

[11] Benchetrit G. Breathing pattern in humans: diversity and individuality[J]. Respiration physiology, 2000, 122(2-3): 123-129.

[12] Yuan G, Drost N A, McIvor R A. Respiratory rate and breathing pattern[J]. McMaster Univ. Med. J, 2013, 10(1): 23-25.

[13] Hudgel D W, Martin R J, Johnson B, et al. Mechanics of the respiratory system and breathing pattern during sleep in normal humans[J]. Journal of Applied Physiology, 1984, 56(1): 133-137.

[14] Obstructive Sleep Apnea,Available at, https://aasm.org/resources/factsheets/sleepapnea.pdf, Accessed in 2023.02.17

[15] Pevernagie D A, Gnidovec-Strazisar B, Grote L, et al. On the rise and fall of the apnea– hypopnea index: A historical review and critical appraisal[J]. Journal of sleep research, 2020, 29(4): e13066.

[16] Iber C. Are we ready to define central hypopneas?[J]. Sleep, 2013, 36(3): 305.

[17] Gruber P, Kwiatkowski T, Flaster E, et al. “Time to equilibration of oxygen saturation using pulse oximetry”, Academic Emergency Medicine, Vol.2,no.9, pp.810-815,1995.

[18] Arnardottir E S, Gislason T. “Quantifying airflow limitation and snoring during sleep”, Sleep Medicine Clinics, Vol.11,no.4, pp.421-434,2016

[19] Shen F, Cheng S, Li Z, et al. “Detection of snore from OSAHS patients based on deep learning”, Journal of Healthcare Engineering, 2020

[20] Sebastian A, Cistulli P A, Cohen G, et al. Automatic Classification of OSA related Snoring Signals from Nocturnal Audio Recordings[J]. arXiv preprint arXiv:2102.12829, 2021.

[21] Tanyer S G, Ozer H. Voice activity detection in nonstationary noise[J]. IEEE Transactions on speech and audio processing, 2000, 8(4): 478-482.

[22] Graf S, Herbig T, Buck M, et al. Features for voice activity detection: a comparative analysis[J]. EURASIP Journal on Advances in Signal Processing, 2015, 2015(1): 1-15.

Chapter 4

Apnea-Hypopnea Index calculation

4.1 Introduction of AHI

The Apnea–Hypopnea Index (AHI) is the number of apnea and hypopnea events per hour during sleep. It is used to indicate sleep apnea severity. The AHI is calculated by dividing the number of apnea and hypopnea events by the number of hours of sleep. It can also separate as Apnea Index and Hypopnea Index[1]. The AHI values for adults are categorized into four levels in the clinic. Normal: AHI<5, Mild sleep apnea: 5≤AHI<15, Moderate sleep apnea: 15≤AHI<30, severe sleep apnea: AHI≥30[2]. AHI is the parameter used most commonly to diagnose and categorize the disease severity of OSA, and the use of the AHI is a common practice in most studies[3-4]. The relationship between AHI levels and health risk is shown in Figure 4.1.

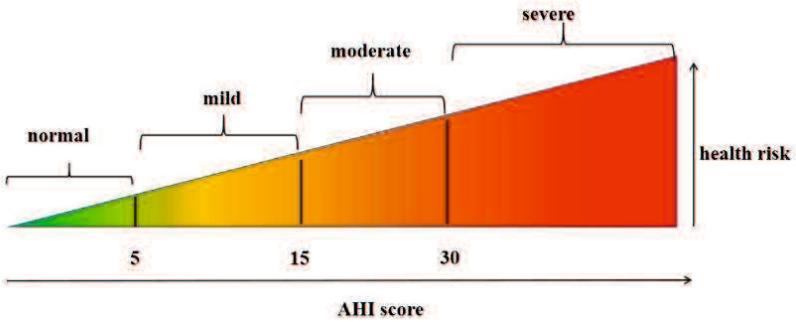

Figure 4.1 AHI level and health risk

The calculation of the Hypopnea Index is more complicated compared with the calculation of the Apnea Index. Most experts agree on the standard definition of apnea as a breathing pause of more than 10 seconds. Hypopneas are more subjective since they occur when upper airways partially collapse. As a result, there are multiple criteria for what counts as hypopnea. Experts have experimented with defining hypopneas according to a certain percentage of decreased airflow, coupled with associated changes in blood oxygen levels or arousals from sleep. However, hypopneas may be measured differently because the American Academy of Sleep Medicine (AASM) has published several versions of the hypopnea definition, and as a result, different definitions of hypopnea can lead to different AHI scores[5-7].

As the apnea and hypopnea clip were identified in chapter 3, to evaluate the accuracy and effectiveness of the identification result, the Apean Index and Hypopnea Index are calculated from identification results and compared with the label extracted from PSG-audio dataset.

4.2 Apnea Index calculation

For each apnea clip identified in chapter 3, the length of the clip is 30 seconds, and the apnea definition is a breathing pause time of more than 10 seconds. For the apnea, at the end of the obstruction, the closed upper airway is suddenly opened, and the pressures of the upper and lower airflows are suddenly balanced. This cause the upper airway to repeat multiple openings and closings in a short period, producing a snoring sound called apneic snoring. This ventilation process with snoring may repeat 2-5 times[8]. Therefore, there is usually one apnea event in each clip. Figure 4.2 shows a 5 minutes length of PSG-audio data with apnea.

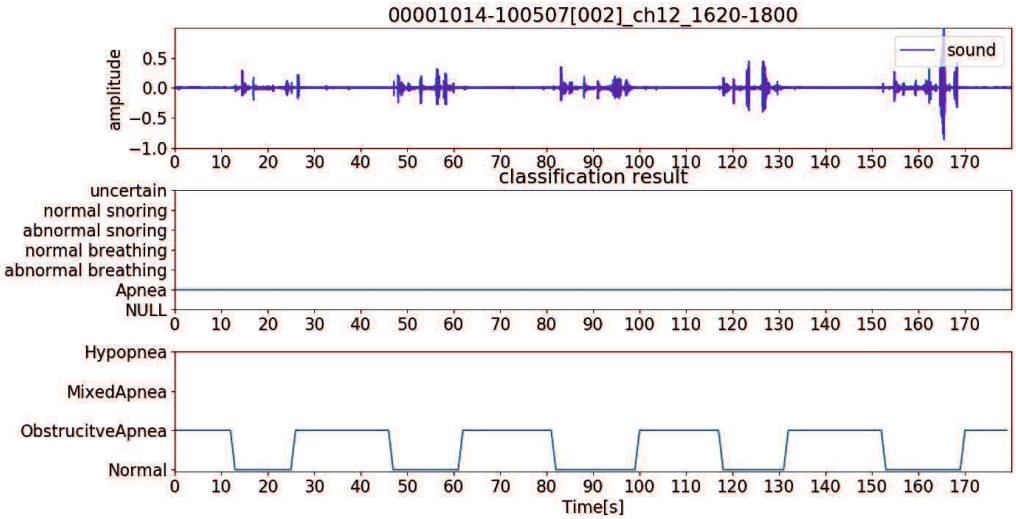

Figure 4.2 a 5 minutes length of PSG-audio data

Although apnea can be identified by breathing pause time, a few clips are still misjudged as hypopnea. These apnea clips usually with pause time around the threshold(10 seconds). Figure 4.3 shows a clip of the pause time just almost 9 seconds, merely more than the threshold. From the clinical aspect, these misjudgments do not change the sum number of AHI. From the breathing quality aspect, mild apnea and severe hypopnea similarly influence sleep quality. Therefore, these misjudgments are acceptable.

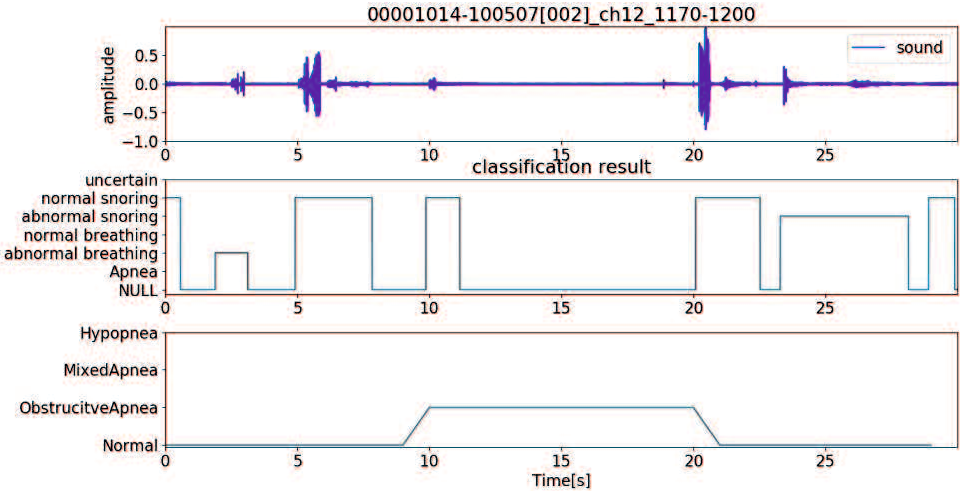

Figure 4.3 apnea clip misjudged as hypopnea

4.3 Hypopnea index calculation

The definition of the American Academy of Sleep Medicine (AASM) is an abnormal respiratory event lasting more than 10 seconds with more than a 30% reduction in thoracoabdominal movement or airflow and more than 4% oxygen desaturation. This is the approved hypopnea definition by the Centers for Medicare and Medicaid Services[9]. Nevertheless, in 2005 the AASM reported that several clinical definitions of hypopnea are in clinical use, and there needs to be a clear consensus[10]. Besides the definition from AASM, in a further attempt to improve standardization, the AASM recently published the manual for the Scoring of Sleep and Associated Events. In this manual, there is a "recommended" and an "alternative" hypopnea definition; and either can be used at the discretion of the clinician or investigator[11]. Besides the AASM definition, other institutions also published similar criteria, such as the Chicago Creterion[12-16].

As hypopnea identification has been discussed in chapter 3 based on two criteria, it will not be discussed here again. However, recent studies suggested that simple snoring also is related to Hypopnea. According to Victor Hoffstein's research, simple snoring does not cause a sustained deterioration of MnO2(mean nocturnal oxygen saturation) but causes the variability of LoO2(lowest nocturnal oxygen saturation) significantly[17]. Based on this research, the airflow during simple snoring is similar to normal respiration. However, after a specific duration, the nocturnal oxygen saturation fluctuation increases and decreases ventilation quality at a moderate level. Although the accurate drop time is not precise, according to the research by Gruber, the interval to equilibration of oxygen saturation is within 4.5 minutes[18]. Therefore after a specific time of successive simple snoring, the breathing state deteriorates and may lead to Hypopnea. The last time to identify this type of Hypopnea is challenging as it probably has an individual variance. However, the last time is more

than 5 minutes, meaning that when normal breathing ends and simple snoring starts, after approximately 5 minutes, the airflow drops to a medium level with high probability and may cause Hypopnea. A type of this clip is shown in Figure 4.4. There are eight normal snoring cycles in this clip, and these cycles have nearly regular intensity and duration. Based on these parameters, it looks like the breathing is in a normal state. However, as it is a clip in the time domain that was snoring last several minutes, the airflow still dropped into the hypopnea range. Based on our research, this type of Hypopnea is rare in the PSG-audio dataset. However, the relationship between simple snoring and Hypopnea still needs more analysis in the future.

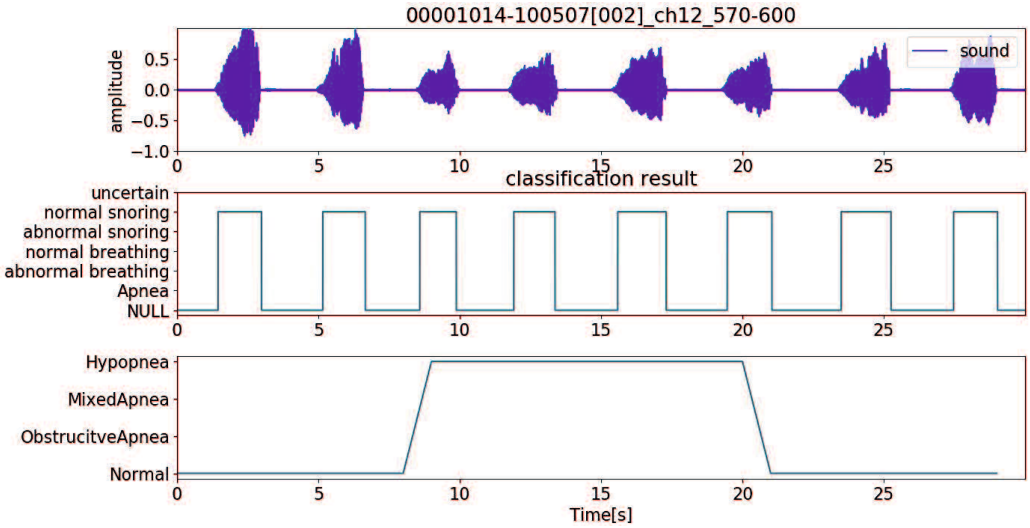

Figure 4.4 normal snoring clip after the normal snoring started more than 5 minutes

4.4 Algorithm performance and robustness evaluation on PSG-audio dataset

Six patients with different Apean-Hypopnea Index(AHI) were selected to test the effectiveness and robustness of the apnea/hypopnea identification method. The accuracy is calculated by Equation 4-1. The characteristic of the data chosen and algorithm performance is shown in table 4-1. The AHI calculation result of two hour recording is shown in Figure 4.5. The first sub-figure is the tracheal sound. The second sub-figure is the nasal pressure signal. The third sub-figure is the apnea/hypopnea events label extracted from the dataset. The fourth sub-figure is the apnea/hypopnea identification result for each clip(the clip length is 30 seconds). It can be seen from the Figure that Apnea Index accuracy is higher than Hypopnea Index.

$$\text{Accuracy} = \frac{\text{correct prediction clip number}}{\text{AHI}} \tag{4.1}$$

Table 4-1 algorithm accuracy on data with different AHI

Patient id	AHI from dataset	AHI from calculation	Algorithm accuracy
00000995-100507-001-002	17	20	71%
00000995-100507-001-001	65	77	77%
00001000-100507-001-001	26	33	69%
00001008-100507-001-001	8	10	75%
00001010-100507-001-001	25	30	80%
00001014-100507-002-002	47	35	74%

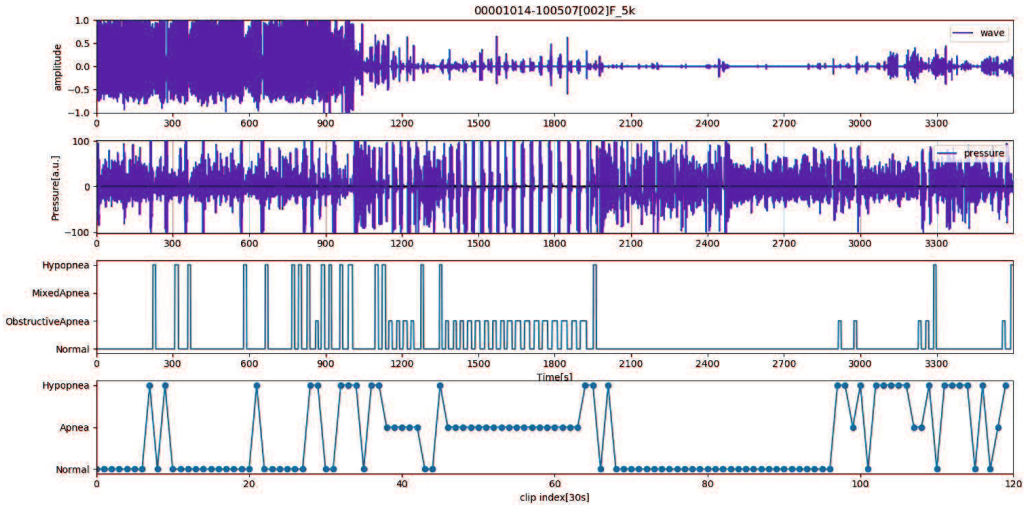

Figure 4.5 AHI calculation result of two hour recording

4.5 Summary

The AHI calculation method based on the breath states of clips is proposed in this chapter. Apnea detection is based on the breathing pause time. However, hypopnea detection is more complicated. 3 criterion is defined to associate the breathing sound with hypopnea. Six patients with different Apean-Hypopnea Index(AHI) were selected to test the effectiveness and robustness of the proposed method. The experiment results show that the average accuracy is more than 70%.

Reference

[1] Berry R B, Brooks R, Gamaldo C E, et al. The AASM manual for the scoring of sleep and associated events[J]. Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine, 2012, 176: 2012.

[2] Hudgel D W. Sleep apnea severity classification—revisited[J]. Sleep, 2016, 39(5): 1165-1166.

[3] Cao W, Luo J, Xiao Y. A review of current tools used for evaluating the severity of obstructive sleep apnea[J]. Nature and Science of Sleep, 2020, 12: 1023.

[4] Pevernagie D A, Gnidovec-Strazisar B, Grote L, et al. On the rise and fall of the apnea− hypopnea index: A historical review and critical appraisal[J]. Journal of sleep research, 2020, 29(4): e13066.

[5] Sands S A, Edwards B A, Terrill P I, et al. Identifying obstructive sleep apnoea patients responsive to supplemental oxygen therapy[J]. European Respiratory Journal, 2018, 52(3).

[6] Eiseman N A, Westover M B, Mietus J E, et al. Classification algorithms for predicting sleepiness and sleep apnea severity[J]. Journal of sleep research, 2012, 21(1): 101-112.

[7] Sateia M J. International classification of sleep disorders[J]. Chest, 2014, 146(5): 1387-1394.

[8] Yang Y, Qin Y, Haung W, et al. Acoustic characteristics of snoring sound in patients with obstructive sleep apnea hypopnea syndrome[J]. Lin Chuang er bi yan hou tou Jing wai ke za zhi= Journal of Clinical Otorhinolaryngology, Head, and Neck Surgery, 2012, 26(8): 360-363.

[9] Korompili G, Kokkalas L, Mitilineos S A, et al. Detecting Apnea/Hypopnea Events Time Location from Sound Recordings for Patients with Severe or Moderate Sleep Apnea Syndrome[J]. Applied Sciences, 2021, 11(15): 6888.

[10] Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005;28(4):499-521

[11] Iber C, Ancoli-Israel S, Chesson A, Quan S; for the American Academy of Sleep Medicine. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007

[12] Ruehland W R, Rochford P D, O'Donoghue F J, et al. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index[J]. sleep, 2009, 32(2): 150-157.

[13] Centers for Medicare and Medicaid Services. National Coverage Determination for Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA). NCD #240.4. 2005 [cited 2008 22 September. Available from: http://www.cms.hhs.gov

[14] Miano S, Paolino M C, Castaldo R, et al. Visual scoring of sleep: A comparison between the Rechtschaffen and Kales criteria and the American Academy of Sleep Medicine criteria in a pediatric population with obstructive sleep apnea syndrome[J]. Clinical neurophysiology, 2010, 121(1): 39-42.

[15] Penzel T, Sabil A K. The use of tracheal sounds for the diagnosis of sleep apnoea[J]. Breathe, 2017, 13(2): e37-e45.

[16] Glos M, Sabil A K, Jelavic K S, et al. Tracheal sound analysis for detection of sleep disordered breathing[J]. Somnologie, 2019, 23(2): 80-85.

[17] Hoffstein V. "Snoring and nocturnal oxygenation: is there a relationship?". Chest, Vol.108,no.2, pp.370-374,1995

[18] Gruber P, Kwiatkowski T, Flaster E, et al. “Time to equilibration of oxygen saturation using pulse oximetry”, Academic Emergency Medicine, Vol.2,no.9, pp.810-815,1995

Chapter 5

Tidal volume estimation based on breathing sound

5.1 Introduction to tidal volume

Tidal volume is the amount of air that moves in or out of the lungs with each respiratory cycle at rest. The tidal volume of an ordinary person is almost 6-10 times the weight in kilograms. Therefore, for a person weights 60 kilograms, the tidal volume is approximately 360-600mL. The average tidal volume is around 500 mL in a healthy adult male and about 400 mL in a healthy female[1]. Tidal volume is one of the parameters for monitoring respiratory ventilation and pulmonary function. It is related to age, gender, respiratory pathology, and body metabolism. When a person breathes in, oxygen from the surrounding atmosphere enters the lungs, then diffuses across the alveolar-capillary interface to reach arterial blood. At the same time, carbon dioxide continuously forms as long as metabolism takes place. Expiration occurs to expel carbon dioxide and prevent it from accumulating in the body. The representation of tidal volume is shown in Figure 5.1. The volume of inspired and expired air that helps keep oxygen and carbon dioxide levels stable in the blood is what physiology refers to as tidal volume[2]. Tidal volume results from inspiration time(in seconds) and inspiration airflow rate(in ml/s). Another concept related to tidal volume is Minute Volume(also known as Respiratory Minute Volume). It is the total amount of air moved into and out of the respiratory system per minute[3]. It is the product of respiratory rate and tidal volume. Therefore the relationship between minute volume, tidal volume, respiratory rate, and airflow rate can be described in Equation 5.1

$$\text{tidal volume} = \text{inspiratory time} * \text{inspiratory airflow rate}$$
$$\text{minute volume} = \text{respiratory rate} * \text{tidal volume} \quad (5.1)$$

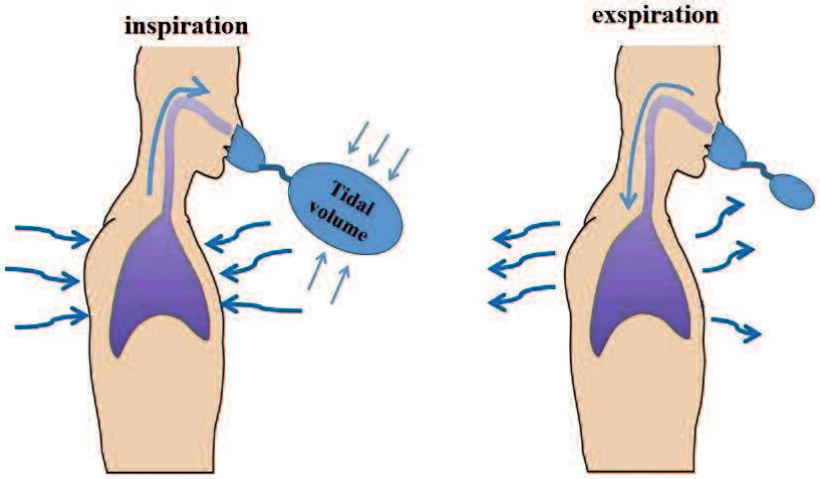

Figure 5.1 the represent of tidal volume

5.2 tidal volume estimation

Many researchers have focused on analyzing the relationship between breathing sound and tidal volume due to its potential to assess breathing quality and lung function. One of the standard models to simulate the upper airway function is the Starling Resistor Model(SRM). This method models the upper airway as a rigid tube with a collapsible segment. Upper (upstream, nasal) and lower (downstream, hypopharyngeal) segments have fixed diameters and defined resistances. According to this model, the airflow rate and breathing sound amplitude follow a relationship compatible with jet noise production in a pipeline[4-5]. This model can be used to predict the effects of transmural pressure on airflow dynamics and the severity of upper airway obstruction during sleep. The representation of SRM is shown in Figure 5.2. Some researchers focused on estimating the respiratory airflow rate through respiratory sounds and proposed various models or algorithms. It is well known that breathing sound amplitude is positively correlated with airflow rate. However, determining the exact quantitative relationship is still challenging. Gavriely proposed that the Breath-Sound Amplitude (BAS) and flow(F) can be expressed in the form of BAS=k*Fα, α is approximately 1.75, significantly less than the second power that some previous research presented[6]. Yap proposed a method to use average power and an exponential model to estimate respiratory flow through tracheal sound, which reached an estimation error of 5.8±3.0%[7]. Natasa used the Blanket Fractal Dimension(BFD) to assess the tidal volume from tracheal sounds recorded by an Android smartphone, the smallest normalized root-mean-squared error of 15.877%±9.246% was obtained with the BFD and exponential model[8]. Yadollahi extracted average power, the logarithm of the variance, and the logarithm of the envelope of tracheal sound as a feature. They compared the ability of these features to fit the flow-sound relationship, suggesting that the logarithm of the variance is the best feature to describe the flow-sound relationship with a linear model[9]. Other studies indicated that the Shannon entropy and sound variance also have an exponential relationship with the respiratory flow[10-11].

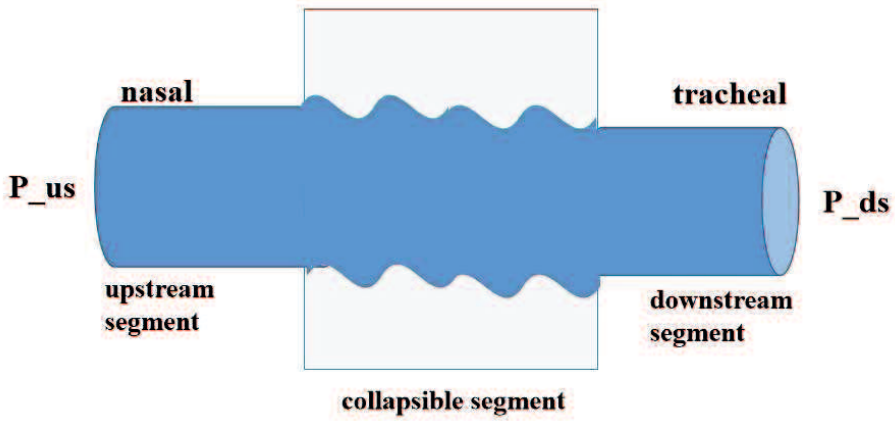

Figure 5.2 the Starling Resistor Model

Most of these papers indicate that the airflow rate and respiratory sound amplitude follow a power law based on the SRM. The relationship used to estimate respiratory flow rate can be presented in Equation 5.2:

$$\log F_{est} = C_1 log(E) + C_2 \tag{5.2}$$

Fest is the estimated airflow rate(ml/min), E is the respiratory sound amplitude, and C1 and C2 are the coefficients. C1 and C2 are determined by the human upper airway structure and can be calculated via a few breaths with a known flow rate for each participant. C2 is determined by the length and diameter of the upper airway, C1 is related to the airflow power(α), representing the sound generation mechanism, and C1 is variable during apnea based on the collapsibility degree. The model parameters can be derived from the breaths with known airflow rates and then used with the rest of the breath sounds to estimate airflow. This procedure is called calibration. The representation of calibration is shown in Figure 5.3. Current methods require calibration to determine the model coefficients C1 and C2. Yadollahi found that the parameters of the flow-sound relationship during sleep and wakefulness are different. For monitoring the tidal volume during sleep, the model parameters should be calibrated with respiratory sounds during sleep[12].

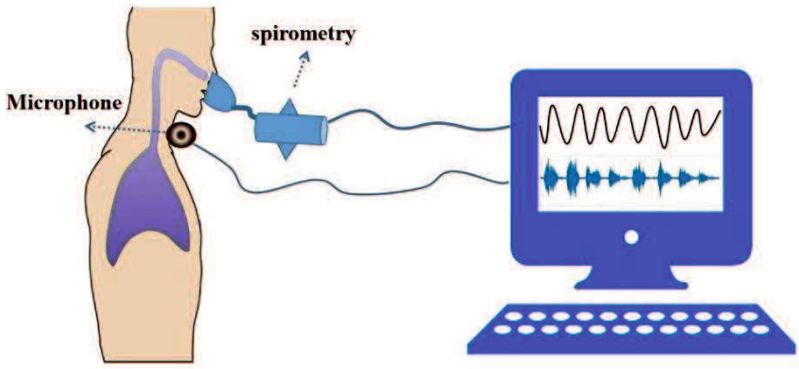

Figure 5.3 the calibration procedure

During sleep, the upper airway of OSA patients is highly variable. Therefore, the coefficients in Equation 5-2 are also highly variable. These above methods often do not apply to predict breathing patterns in the human upper airway during snoring or apnea. During snoring, the sound amplitude is higher than normal breathing. On the contrary, the respiratory airflow is lower than normal breathing. During an apnea, the airflow vanishes and emerges periodically. The main reason is that the upper airway is usually in an unstable state. As the airflow-amplitude relationship is related to upper airway states, these coefficients are also highly variable during snoring or apnea. Based on the abovementioned conclusions, estimating tidal volume using different methods based on the breathing

states is reasonable. In this paper, qualitative and quantitative tidal volume estimation methods are proposed based on breathing states.

5.3 Quantitative tidal volume estimation

To evaluate the calculation accuracy, the nasal cannula pressure is used as the surrogate signal of airflow rate. A nasal cannula is usually placed into the nose, and a pressure transducer is attached. It detects the pressure fluctuations caused by inspiration and expiration. Much research shows that the nasal cannula pressure signal can achieve a breathing monitoring performance comparable to a conventional pneumotachograph[13-15]. Thurnheer proposed that the square-root transform of nasal pressure and airflow follows a linear relationship[16]. Montserrat simultaneously measured nasal flow and pneumotachograph signal on six healthy subjects and verified that the nonlinear square root of nasal pressure could fit the pneumotachograph signal[17]. Therefore the relationship used to estimate respiratory flow rate from nasal pressure can be presented in Equation 5.3

$$F_{est} = k_1\sqrt{P} \tag{5.3}$$

Fest is the estimated airflow rate(ml/min), and P is the nasal pressure. k1 is coefficients. Although the absolute value of Fest needs to calibrate, a relative Fest value without calibration can also use to analyze the airflow rate change trend. The value of log(E) and can be used as the surrogate of airflow. A 15 seconds nasal pressure signal and the corresponding tracheal sound are shown in Figure 5.4. The tracheal sound amplitude is normalized to [-1,1], and the nasal pressure signal is between [-100,100]. During the inspiration, the chest wall and diaphragm expand to increase the alveoli volume so airflow into the lungs. This cause the alveolar pressure to drop to negative. After the mid-inspiration, the intra-alveolar pressure rises to 0. During the expiration, the volume of the lungs decreases, causing the alveolar pressure becomes positive[18]. The area under the pressure signal curve corresponds to air volume breathing in or out.

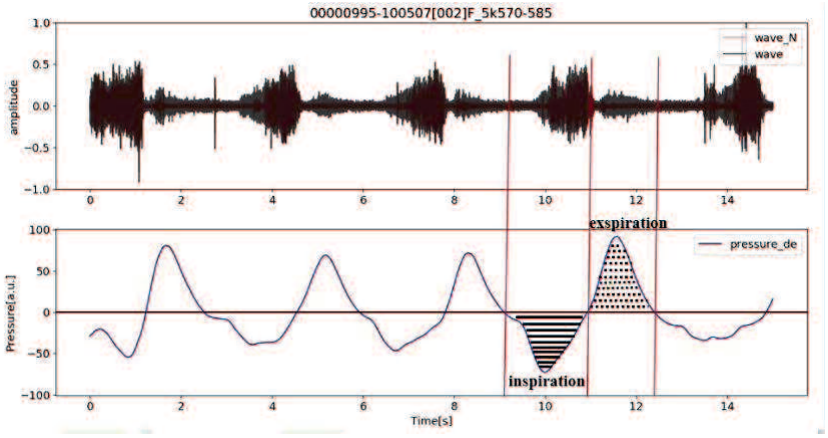

Figure 5.4 tracheal sound(top) and nasal cannula pressure(bottom) signal

5.3.1 normal breathing tidal volume estimation

During normal breathing state, the upper airway keeps stable. The breathing sound amplitude follows a logarithm relationship described in equations 5-3. Once the model coefficients C1 and C2 are calibrated, the airflow rate can be calculated in absolute value(ml/min). The tidal volume(TV) can be calculated by integration of airflow rate with inspiration time.

$$TV = C_1 \int F_{est}(t)\, dt \tag{5.4}$$

As the envelope values are positive, negative nasal pressure signals (inspiration) are converted into positive values. Only the inspiration part of the breathing cycle is discussed, as the inspiration part of tracheal sound has a higher signal intensity than expiration. The relative tidal volume calculation based on tracheal sound is as follows:

(1) calculate the envelope of tracheal sound by TCW.

(2) Segment the inspiration and expiration of breathing sound.

(3) Calculate the inspiration airflow rate by equation 5-2 based on envelope of inspiration

(4) Calculate the inspiration airflow rate by equation 5-2 based on the nasal pressure, as a reference for evaluation.

The result of one clip with normal breathing state is shown in Figure 5.5. The sub-figure from top to bottom is (a)tracheal sound, (b)nasal pressure, (c)relative tidal volume calculated by breathing sound(TiV_sound), (d)relative tidal volume calculated by nasal cannula pressure(Tiv_pressure). The Pearson correlation coefficient is introduced as a metric to evaluate the correlation of TiV_sound and TiV_pressure. The Pearson correlation coefficient is 0.68, which shows a high correlation.

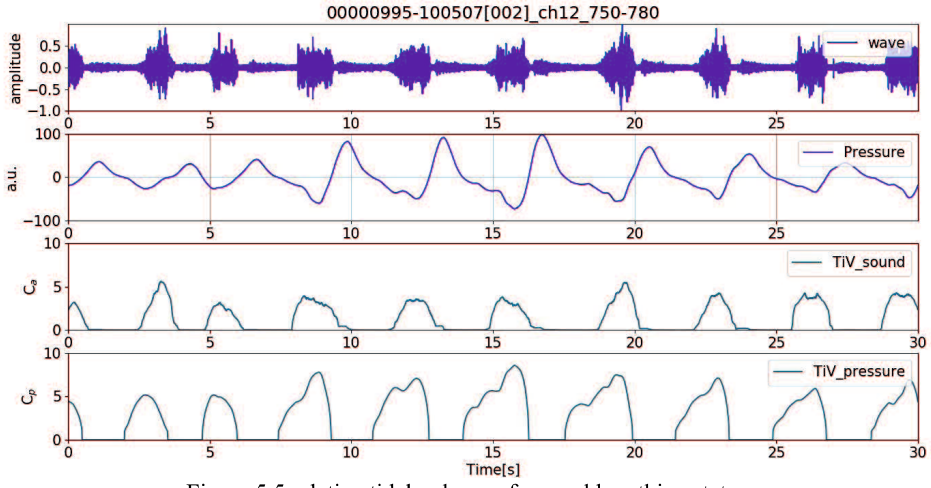

Figure 5.5 relative tidal volume of normal breathing state

5.3.2 Normal snoring tidal volume estimation

During the normal snoring state, the upper airway slightly or partially collapses. According to the AASM, simple snoring usually does not interfere with the patient's sleep or cause excessive daytime sleepiness[19]. Simple snoring is generally not considered a health threat compared with apnea. However, it is a probable risk factor for the development of OSA.

As the upper airway can be regarded as a stable state, the normal snoring sound amplitude-flow can also be regarded as follows a logarithm relationship described in equation 5-1 but with a different model coefficient compared with normal breathing. According to the research by Saha, snoring sound generation is associated with upper airway diameter, length, and wall thickness. Upper airway narrowing is the only factor that positively and significantly contributes to snoring intensity[20]. The upper airway partially collapses, reducing the diameter. Meanwhile, the length and wall thickness are invariable. Therefore the coefficients of normal snoring can be regarded as proportional to the coefficients of normal breathing. Based on our research, the practical value can be set as 0.4. The result of one clip with a normal snoring state is shown in Figure 5.6. The sound intensity is much higher than normal breathing, while the pressure signal drops significantly to a lower level.

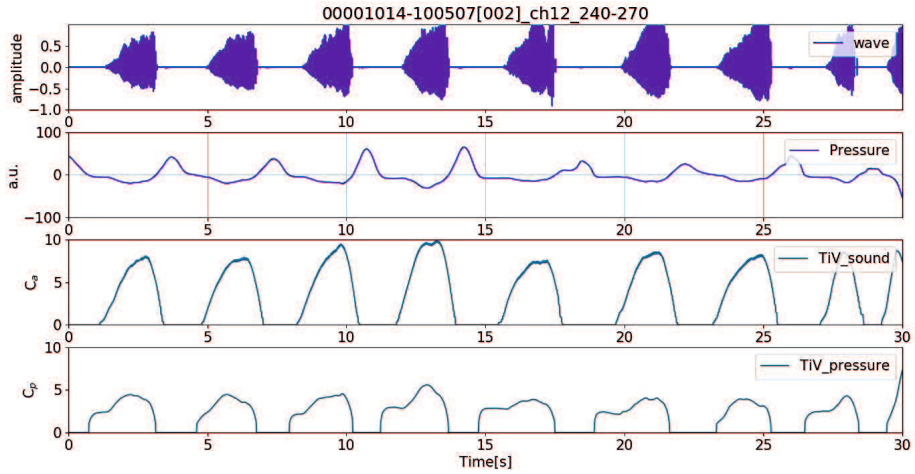

Figure 5.6 tidal volume of normal snoring state

5.4 Qualitative tidal volume estimation

5.4.1 tidal volume estimation during abnormal breathing and abnormal snoring

As mentioned in chapter 3, the upper airway anatomy is unstable during abnormal breathing and abnormal snoring. Therefore, the tidal volume can not be calculated quantitatively. However, it is possible to estimate the tidal volume qualitatively. The SpO2 is a reading that shows the amount of oxygen available in human blood to deliver to the heart, brain, lungs, and other muscles and organs. The LoO2(lowest nocturnal oxygen saturation) is the lowest SpO2 value during a specific time and has a high correlation with tidal volume. The LoO2 is usually divided into three levels in the clinic. Large

than 95% is considered a high level, less than 90% is considered low(hypoxemia), and between 95% and 90% is considered medium(mild) hypoxemia[21-22]. Therefore, the tidal volume levels are divided into high/medium/low levels corresponding to the LoO2 levels.

The tidal volume level is mainly determined by airflow intensity and breathing rate. Based on the definition of abnormal breathing and abnormal snoring states in chapter 3, the airflow intensity and breathing rate are above the apnea threshold. Therefore the tidal volume levels are at either average or median levels. Although the airflow intensity could not be calculated quantitatively from breathing sounds, the breathing rate calculated is relatively accurate. The criterion defined in chapter 3 for identifying hypopnea is used to separate the tidal volume levels into normal/median levels. An abnormal breathing clip with tidal volume at a normal level is shown in Figure 5.7. The sound and pressure signals show that the breathing state is unstable, and the breathing intensity and cycle duration are irregular. However, from the ventilation aspect, the tidal volume is normal as the breathing rate is in the normal range. An abnormal breathing clip with tidal volume in the median level is shown in Figure 5.8. The breathing rate is below the normal range.

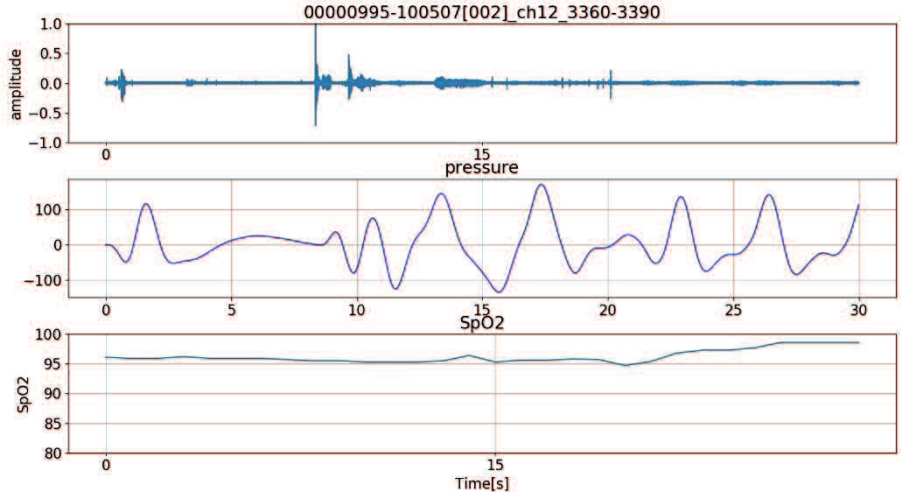

Figure 5.7 abnormal breathing with normal tidal volume level

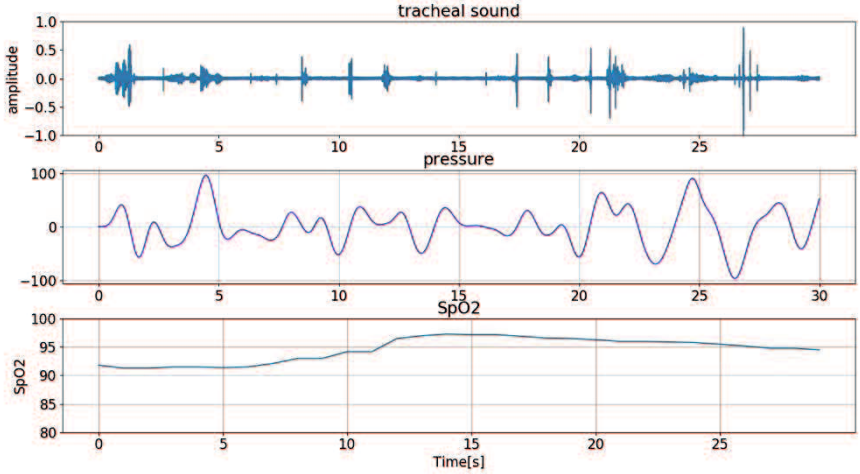

Figure 5.8 abnormal breathing with median tidal volume level

5.4.2 tidal volume estimation during apnea and hypopnea

During apnea or hypopnea, the upper airway is highly variable, and the state of obstruction, collapse, and opening occur repeatedly. It is challenging to estimate the tidal volume value quantitatively. However, it is possible to assess the tidal volume level qualitatively. Based on the research by Ma[23], nocturnal hypoxemia severity is proportional to the breathing pause time. To evaluate the severity of apnea and hypopnea, the Maximum Breathing Pause Interval(MBPI) is calculated as a parameter. According to the Apnea definition, the threshold to distinguish the low/medium level of tidal volume is set to 15 seconds. Clips with MBPI<15 and MBPI>15 are shown in Figure 5.9 and Figure 5.10, respectively. The SpO2 change lags with the breath airflow. It starts to drop after a short time of limited airflow. Since the clip duration is set as 30 seconds, it is long enough to ignore the lagging effect.

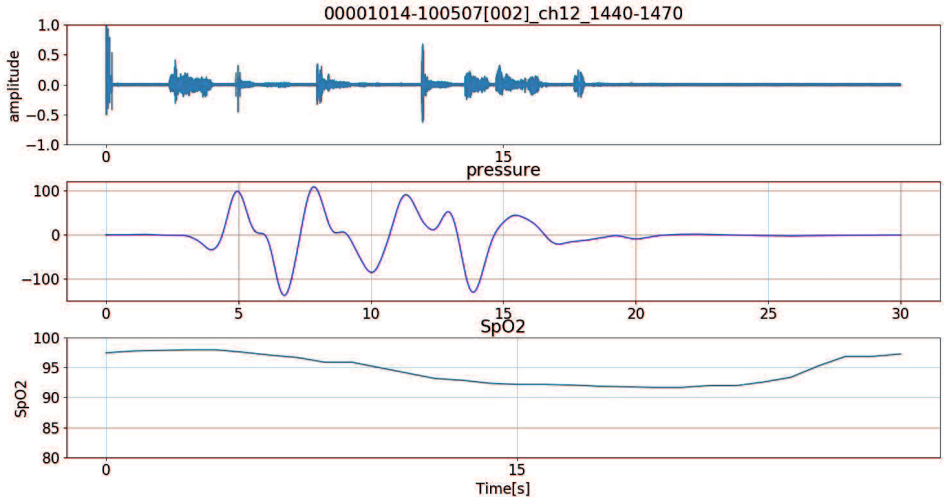

Figure 5.9 apnea with MBPI<=15,LoO2 between 90%-95%

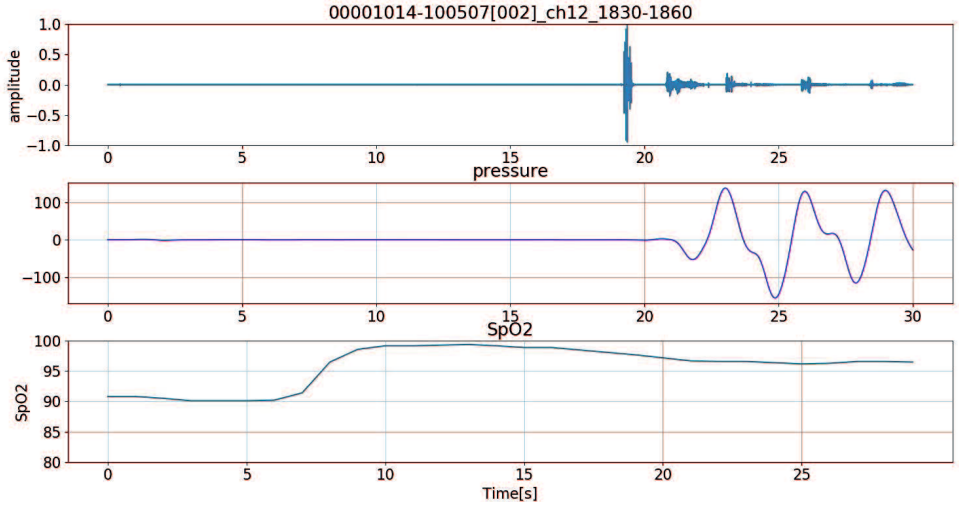

Figure 5.10 apnea with MBPI>15,LoO2 between below 90%

The tidal volume level estimation during hypopnea is challenging as it is related to the last time. According to Victor Hoffstein's research, the upper airway collapsing does not cause a sustained deterioration of MnO2(mean nocturnal oxygen saturation) but causes significant variability of LoO2(lowest nocturnal oxygen saturation)[24]. Based on this research, the tidal volume level during hypopnea beginning is similar to normal respiration. Still, after a certain duration, the nocturnal oxygen saturation fluctuation increases and decreases ventilation quality at a moderate level. Although the accurate SpO2 drop time is not clear, according to the research by Gruber, the interval to equilibration of oxygen saturation is within 4.5 minutes[25]. Therefore the SpO2 drop threshold is set at 4 minutes, meaning that when normal breathing ends and hypopnea starts, after approximately 4 minutes, the SpO2 drops to a low level with high probability. Figure 5.11 is hypopnea that lasts less than 4 minutes. The SpO2 drops slightly, and the LoO2 drops to between 95% and 90%. Figure 5.12 is hypopnea that lasts more than 4 minutes. The SpO2 drops slightly, and LoO2 drops to below 90%.

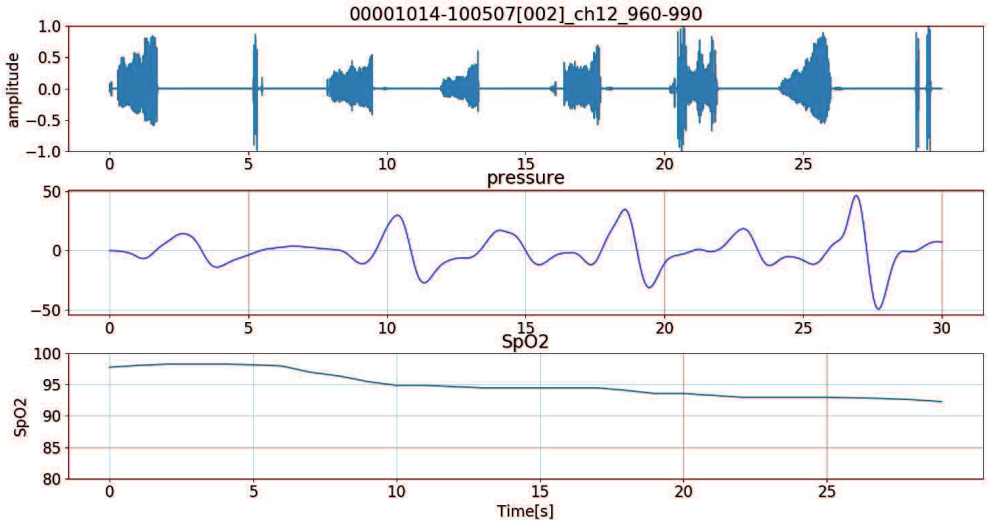

Figure 5.11 hypopnea lasts less than 4 minutes from hypopnea starts

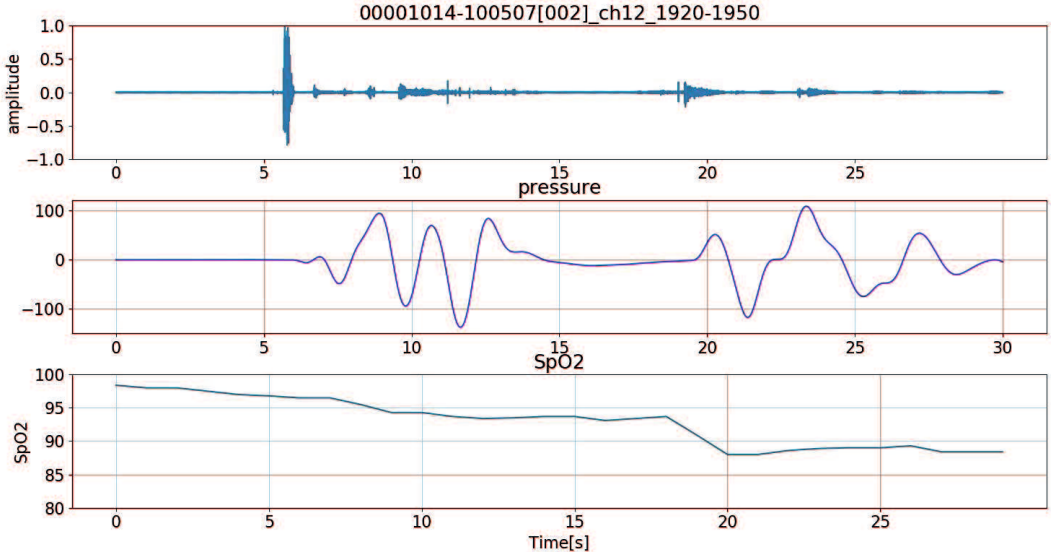

Figure 5.12 hypopnea lasts more than 4 minutes

The overall results of one hour are shown in Figure 5.13. The top sub-figure is the tidal volume level calculated by the proposed method, the x-axis represents the clip index, and each clip is 30 seconds long. The middle sub-figure is the LoO2 which is divided into high/medium/low levels, and the uncertain level corresponds to the uncertain states. The bottom sub-figure is the SpO2 level that is used to calculate the LoO2.

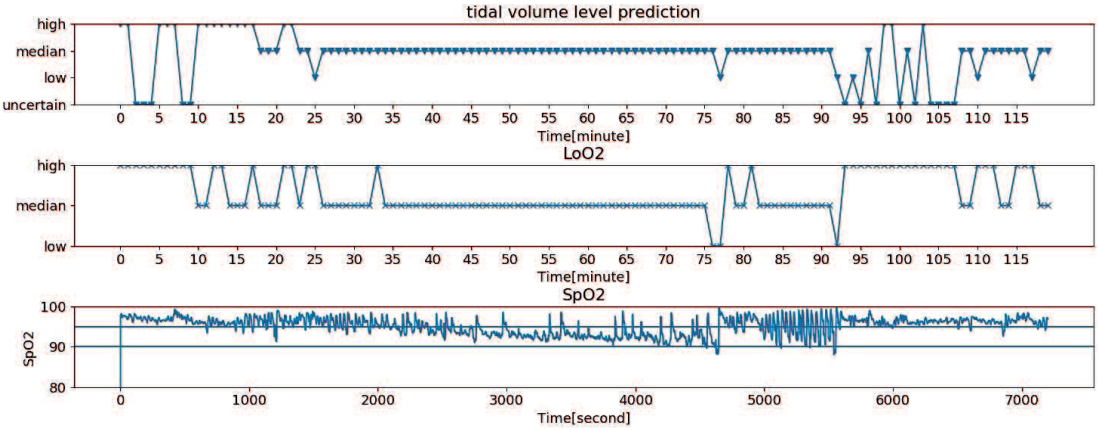

Figure 5.13 The prediction result: (top)tidal volume level prediction, (middle)LoO2, (bottom)SpO2

The accuracy is calculated by Equation 5-5. Six patients with different Apean-Hypopnea Index(AHI) were selected to test the effectiveness and robustness of the proposed method. The characteristic of the data chosen and algorithm performance is shown in table 5-1. The algorithm accuracy is 88.3% in the group with mild apnea. As for the moderate apnea group, the algorithm accuracy slightly drops to 85.8%. In the severe apnea group that sound signal containing ambient noise, the algorithm accuracy is still above 83%.

$$\text{Accuracy} = \frac{\text{correct prediction number}}{\text{total number} - \text{uncertain number}} \tag{5.5}$$

Table 5-1 algorithm accuracy on data with different characteristics

Apnea severity	Patient number	Data length(hour)	Data characteristic	Algorithm accuracy
mild	1	2	Mild apnea, no simple snoring	88.3%
moderate	3	6	Moderate apnea, little simple snoring	85.8%
Severe	2	4	Severe apnea, little simple snoring, containing ambient noise	83.3%

5.5 Summary

A tidal volume level prediction method is proposed based on unsupervised clustering and snoring parameters. The amplitude-airflow follows a logarithm relationship. The normal breathing and snoring

tidal volume are calculated quantitatively based on this relationship. The abnormal breathing and abnormal snoring tidal volume are calculated qualitatively based on the breathing rate. The apnea and hypopnea tidal volume is calculated qualitatively based on the MBPI and hypopnea last time. Quantitative estimation can provide an absolute airflow value after the coefficients' calibration. The qualitative method can offer a coarse-grained tidal volume level estimation that does not need any calibration. In addition, this method can be used for sleep breathing monitoring in a home environment.

Reference

[1] Malhotra A. Low-tidal-volume ventilation in the acute respiratory distress syndrome[J]. New England Journal of Medicine, 2007, 357(11): 1113-1120.

[2] Ferguson N D, Frutos-Vivar F, Esteban A, et al. Airway pressures, tidal volumes, and mortality in patients with acute respiratory distress syndrome[J]. Critical care medicine, 2005, 33(1): 21-30.

[3] Hallett S, Toro F, Ashurst J V. Physiology, tidal volume[J]. 2018.

[4] Wellman A, Genta P R, Owens R L, et al. Test of the Starling resistor model in the human upper airway during sleep[J]. Journal of applied physiology, 2014, 117(12): 1478-1485.

[5] Schwartz A R, Smith P L. CrossTalk proposal: the human upper airway does behave like a Starling resistor during sleep[J]. The Journal of physiology, 2013, 591(Pt 9): 2229.

[6] Gavriely N, Cugell D W. “Airflow effects on amplitude and spectral content of normal breath sounds”, Journal of applied physiology, Vol.80, no.1, pp.5-13,1996

[7] Yap Y L, Moussavi Z. “Acoustic airflow estimation from tracheal sound power”, IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Vol.2, pp.1073-1076,2002

[8] Reljin N, Reyes B A, Chon K H. “Tidal volume estimation using the blanket fractal dimension of the tracheal sounds acquired by smartphone”, *Sensors*, Vol.15, no.5, pp.9773-9790,2015

[9] Yadollahi A, Moussavi Z. “Comparison of flow-sound relationship for different features of tracheal sound”, 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp.805-808, 2008

[10] Yadollahi A, Moussavi Z M K. A robust method for estimating respiratory flow using tracheal sounds entropy[J]. IEEE Transactions on Biomedical Engineering, 2006, 53(4): 662-668.

[11] Morgenstern C, Jane R, Schwaibold M, et al. Characterization of inspiratory flow limitation during sleep with an exponential model[C]//2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008: 2439-2442.

[12] Yadollahi A, Montazeri A, Azarbarzin A, et al. Respiratory flow–sound relationship during both wakefulness and sleep and its variation in relation to sleep apnea[J]. Annals of biomedical engineering, 2013, 41(3): 537-546.

[13] Norman R G, Ahmed M M, Walsleben J A, et al. Detection of respiratory events during NPSG: nasal cannula/pressure sensor versus thermistor[J]. Sleep, 1997, 20(12): 1175-1184.

[14] Montserrat J M, Farré R. Breathing flow disturbances during sleep: can they be accurately assessed by nasal prongs?[J]. American journal of respiratory and critical care medicine, 2002, 166(3): 259-260.

[15] Fu J, Teng W N, Li W, et al. "Estimation of Respiratory Nasal Pressure and Flow Rate Signals Using Different Respiratory Sound Features", *IRBM*, 2021

[16] Thurnheer R, Xie X, Bloch K E. Accuracy of nasal cannula pressure recordings for assessment of ventilation during sleep[J]. American journal of respiratory and critical care medicine, 2001, 164(10): 1914-1919.

[17] Montserrat J M, Farré R, Ballester E, et al. Evaluation of nasal prongs for estimating nasal flow[J]. American journal of respiratory and critical care medicine, 1997, 155(1): 211-215.

[18] Breathing Cycle http://www.pathwaymedicine.org/breathing-cycle

[19] Counter P, Wilson J A. The management of simple snoring[J]. Sleep Medicine Reviews, 2004, 8(6): 433-441.

[20] Saha S, Bradley T D, Taheri M, et al. A subject-specific acoustic model of the upper airway for snoring sounds generation[J]. Scientific reports, 2016, 6(1): 1-10.

[21] Madan A. Correlation between the levels of SpO2and PaO2[J]. Lung India, 2017, 34(3).

[22] Terragni P P, Rosboch G, Tealdi A, et al. "Tidal hyperinflation during low tidal volume ventilation in acute respiratory distress syndrome", American journal of respiratory and critical care medicine, Vol.175, no.2, pp. 160-166,2007

[23] Ma C, Zhang Y, Liu J, et al. "A novel parameter is better than the AHI to assess nocturnal hypoxaemia and excessive daytime sleepiness in obstructive sleep apnoea", Scientific Reports, Vol.11,no.1,pp.1-8,2021

[24] Hoffstein V. "Snoring and nocturnal oxygenation: is there a relationship?". Chest, Vol.108,no.2, pp.370-374,1995.

[25] Gruber P, Kwiatkowski T, Flaster E, et al. "Time to equilibration of oxygen saturation using pulse oximetry", Academic Emergency Medicine, Vol.2,no.9, pp.810-815,1995

Chapter 6

Application extension

6.1 clustering method applied to heart sound monitoring

6.1.1 Introduction to heart sound monitoring

Studies have shown that Sleep-Related Breathing Disorders are closely related to Cardiovascular Disease(CVD). There is a close association between SRBD and myocardial infarction, congestive heart failure, and atrial fibrillation[1]. People with SRBD, such as OSA or insomnia, also have a higher rate of suffering from heart arrhythmias and coronary artery disease than the general public[2]. Heart sound auscultation is an easy way to analyze CVD compared with Electrocardiography(ECG). Many pathologic cardiac conditions can be diagnosed by analyzing heart sounds. Similar to the breathing sound acquisition, the heart sound recording also often gets contaminated with breathing sounds or ambient noise. The first step of heart sound analysis is to extract analyzable parts from the record file. Much research has been done to classify heart sounds for analysis based on sound quality. Tanveer proposed a shape-based approach to retrieve similar heart sounds. The morphological variations of audio envelopes were extracted and compared with a constrained non-rigid transform[3]. Guy Amit proposed a method for identifying morphological characteristics of heart sounds and classifying them into different states using the hierarchical clustering algorithm[4]. Kamarulafizam proposed a method to discriminate the typical heart sound from the abnormal heart sound based on Time-Frequency Distribution. The classification proceeded with an Artificial Neural Network and achieved a high accuracy[5].

The research mentioned above focuses on extracting various features of heart sounds from the time or frequency domain and uses different algorithms to classify them into different categories. But there are still many problems current research did not resolve:

(1)The dataset used in the abovementioned research is all high-quality heart sounds with little noise or contamination. However, the rise of mobile devices such as home stethoscopes and wearable devices has made it possible to predict and monitor the danger of CVD during sleep in a home environment. Audience noise is often mixed in the sound file when recording heart sounds in a home environment, complicating subsequent analysis[6]. It is challenging to detect the noise parts from the heart sound parts. A heart sound segment contaminated with noise is shown in Figure 6.1.

(2)The recorded heart sounds are often contaminated with lung sounds or snoring, which can also

be helpful in CVD monitoring. Snoring is usually a sign of OSA, which raises the risk of stroke, heart attack, and other cardiovascular problems. Nevertheless, current methods only classify heart sounds as healthy and pathological while ignoring this cardiopulmonary condition. A heart sound segment contaminated with snoring is shown in Figure 6.2.

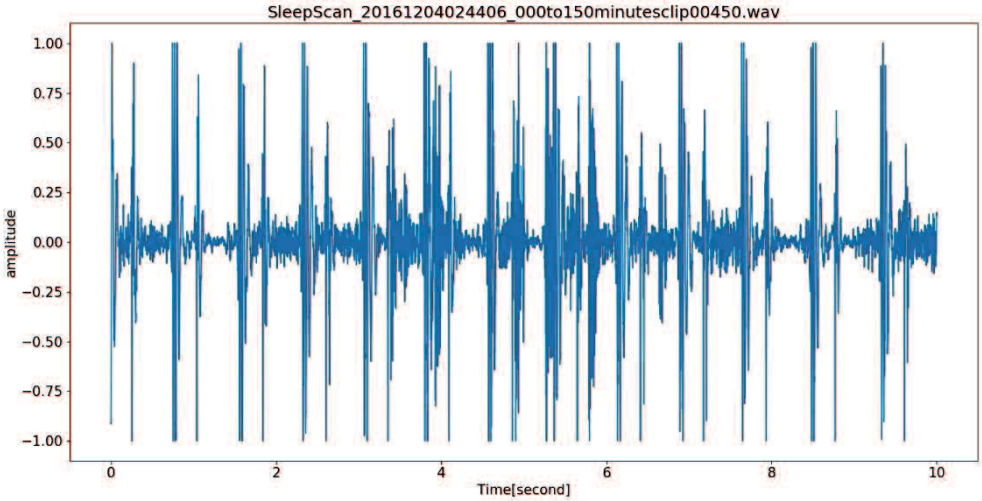

Figure 6.1 heart sound contaminated with noise

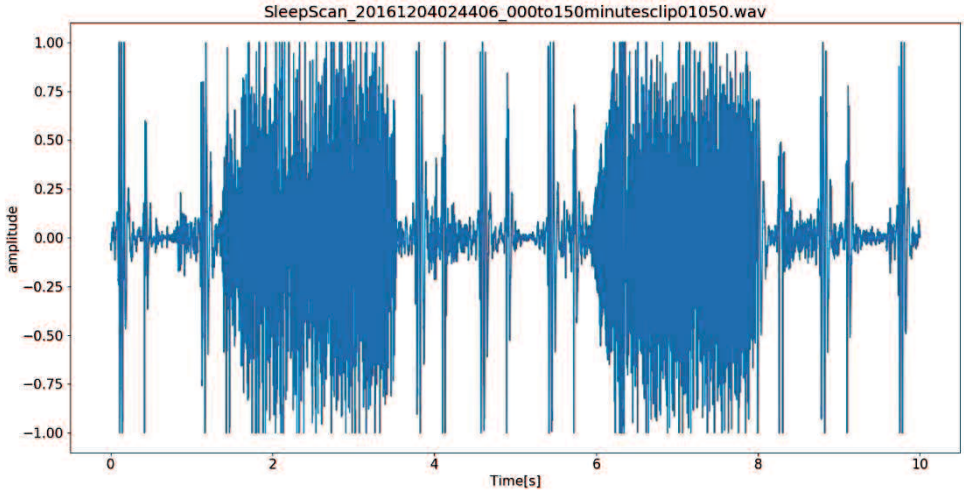

Figure 6.2 heart sound contaminated with snoring sound

The breathing sound clustering method proposed in chapter 2 was adapted and applied to heart sound for analysis. Different analysis methods can proceed on each cluster to monitor CVD danger based on cluster characteristics. Several modifications are made based on the difference between the characteristics of heart sound and breathing sound..

(1) Compared with breathing sound, heart sound energy concentrate in a narrow frequency band in low frequency, mainly in 20-200Hz, therefore, the preprocessing is different from breathing sound.

(2) The number of Mel filter banks used for filtering heart sound could be reduced compared with breathing sound.

6.1.2 Heart sound monitoring method

The proposed method consisted of 5 main steps. The first step is data acquisition. The heart sound is recorded with a chest piece and a digital recorder during sleep. The second step is preprocessing. The audio file was filtered and segmented into clips. The third step is feature extraction. MFCC vectors were extracted as a feature to represent each segmented clip. The fourth step is clustering with AHC. The Euclidean distance was calculated between each clip and formed a distance matrix. The dendrogram is built to show the structure of the audio file. The fifth step is cluster analysis, optimal cluster numbers were set based on the structure of the dendrogram, and the property of each cluster can be determined by the location of the MFCC vector in space. The flowchart of the whole system is illustrated in Figure 6.3.

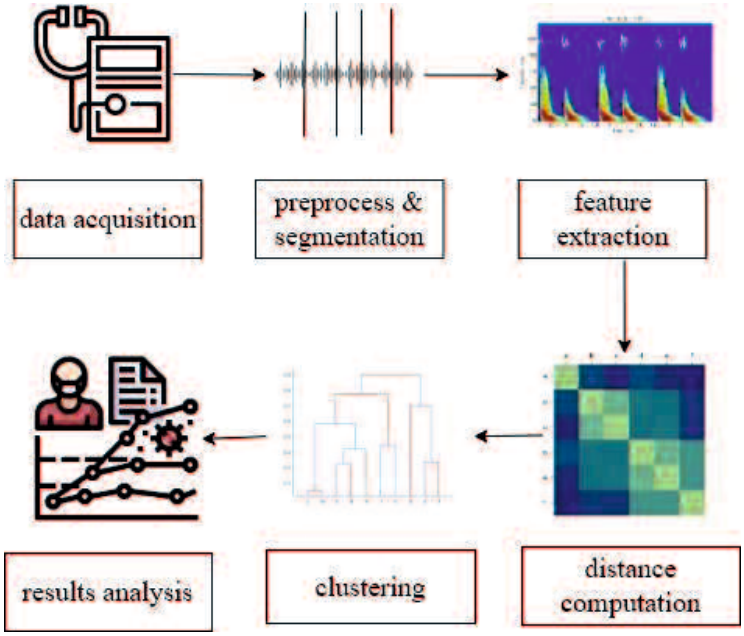

Figure 6.3 Flowchart of heart sound monitoring system

6.1.3 Data Acquisition

The heart sound was recorded by a small chest piece (3M Littmann) attached to the mitral position by adhesive tape. The chest piece was connected with a digital recorder (Olympus Voice-Trek V-843) with wires to record audio. The acquisition proceeded during the night after the subjects had fallen asleep. The record file format is MP3, the sampling rate is 44.1k Hz, and the bit depth is 16-bit. As it is non-intrusive, it does not cause uncomfortable. The devices used in the acquisition are shown in Figure6.4, and the representation of the data acquisition system is shown in Figure 6.5.

Figure6.4 chest piece and digital recorder used in the system

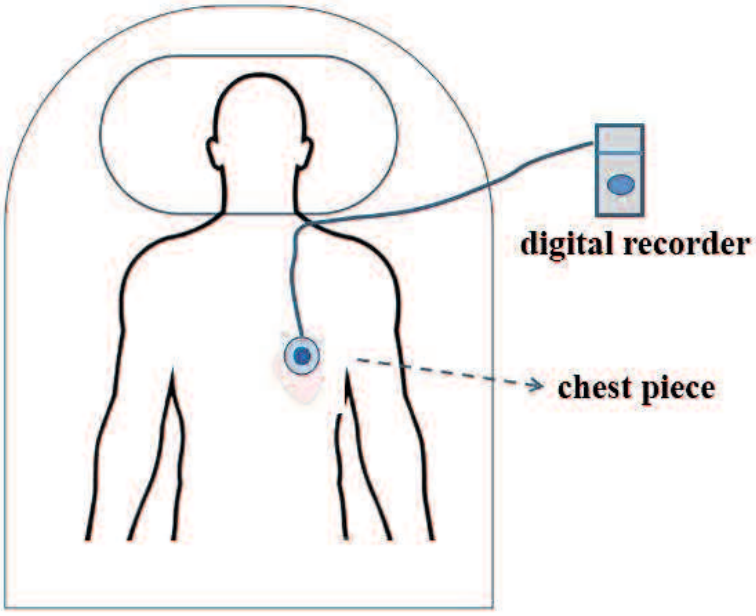

Figure 6.5 Abstract illustration of the heart sound acquisition system

The record file was preprocessed into .wav format. The first step of preprocessing is filtering and denoising. As heart sound frequency is concentrated in 20-200Hz, a 20-200 Hz Butterworth bandpass filter was used to filter noise. The recording files were downsampled to 400Hz. The second step of preprocessing is segmentation. The entire file was cut into clips. The duration of the clip length is settled by considering the micro and the macro aspect. One clip should be short enough to separate the noise or respiration sound from the heart sound. Therefore the audio signal in one clip is stable.

Meanwhile, the clip should be long enough to contain at least several heartbeat cycles. As the heart rate of a healthy person is usually between 60 to 100 beats per minute while resting, the record file was segmented into clips 10 seconds in length. Each clip contains around 10-15 heart sound cycles.

6.1.4 Feature Extraction and Similarity Calculation

Similar to the process steps in chapter 2, the MFCC vectors are extracted as feature vectors. The similarity calculation is also the same. Figure 6.6 shows the waveform and the Mel-spectrum of a heart sound clip with a duration of 10 seconds.

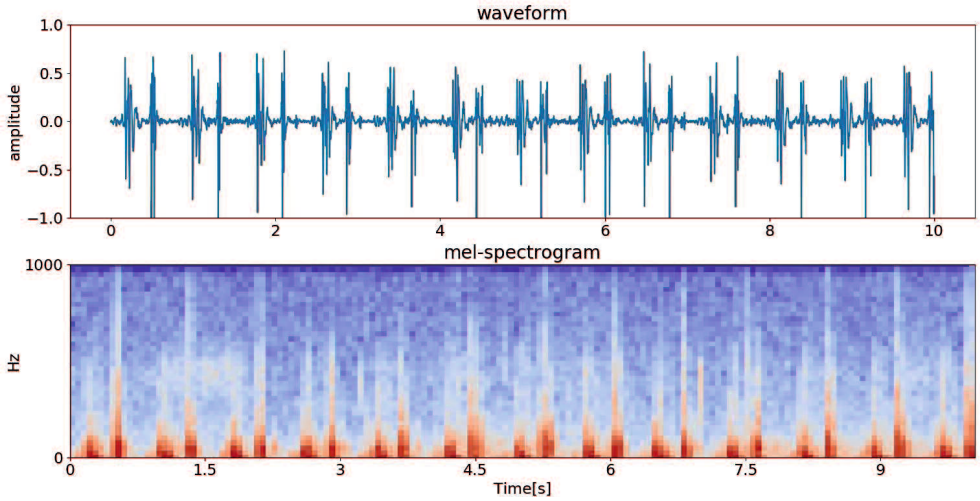

Figure 6.6 The waveform, Mel-spectrum of a heart sound clip

6.1.5.Experiment and Result

Student volunteers were selected to participate in experiments. The recording files were processed with python 3.7. A 173 minutes length file was chosen from all data and segmented into 10 seconds length clips. Therefore 1038 clips were used in the experiments. The STFT spectrum window length is 64ms with an overlap of 32ms. The 20 Mel-scale filters were set in MFCC extraction. As many researchers suggested that the 13-dimension MFCC coefficients used in ASR achieved good performance[10], 13-dimension MFCC coefficients were used in the experiments. The distance matrix size is a symmetry matrix with a size (of 1038,1038). The dendrogram of the clustering result is shown in Figure 6.7. Based on the structure of the dendrogram. The dendrogram was divided into 4 clusters. The cluster1, cluster2, cluster3, and cluster4 were presented with red, green, blue, and cyan, respectively.

The first three dimensions of each MFCC vector are used to visualize the MFCC matrix in a 3-dimensional space. The space is shown in Figure 6.8. The three axes are the zeroth, first, and second coefficients of the MFCC vector. The color of each point corresponds with the dendrogram. As there are 1038 data points, the dendrogram was truncated to show the main structure of all data. One clip was selected from each cluster as the representative, Clips are labelled based on the location in the 3D space. The property and accuracy of each cluster are listed in Table 6.1.

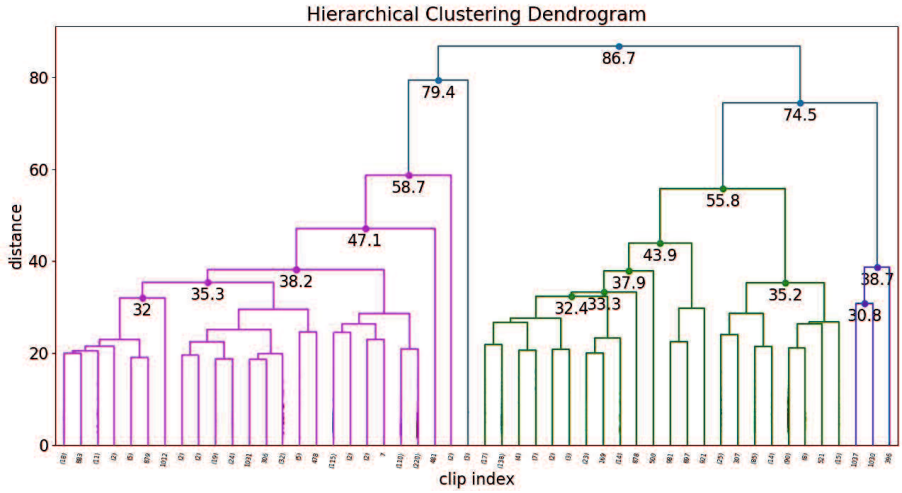

Figure 6.7 The truncated dendrogram of the clustering result

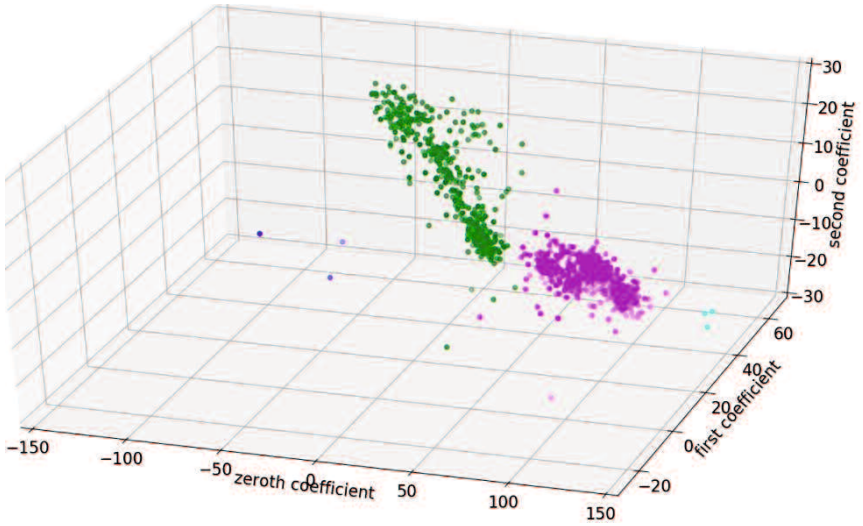

Figure.6.8 The 3D visualization of feature space

Table 6.1 cluster result

Cluster NO.	number	property	color
Cluster 1	3	Heart sound with noise	cyan
Cluster 2	453	Heart sound	green
Cluster 3	3	Ambient noise	blue
Cluster 4	579	Heart sound with lung sound	magenta

The waveform and spectrum of examples present for each cluster were shown in Figure 6.9. Each cluster can be analyzed by the different method based on its property.

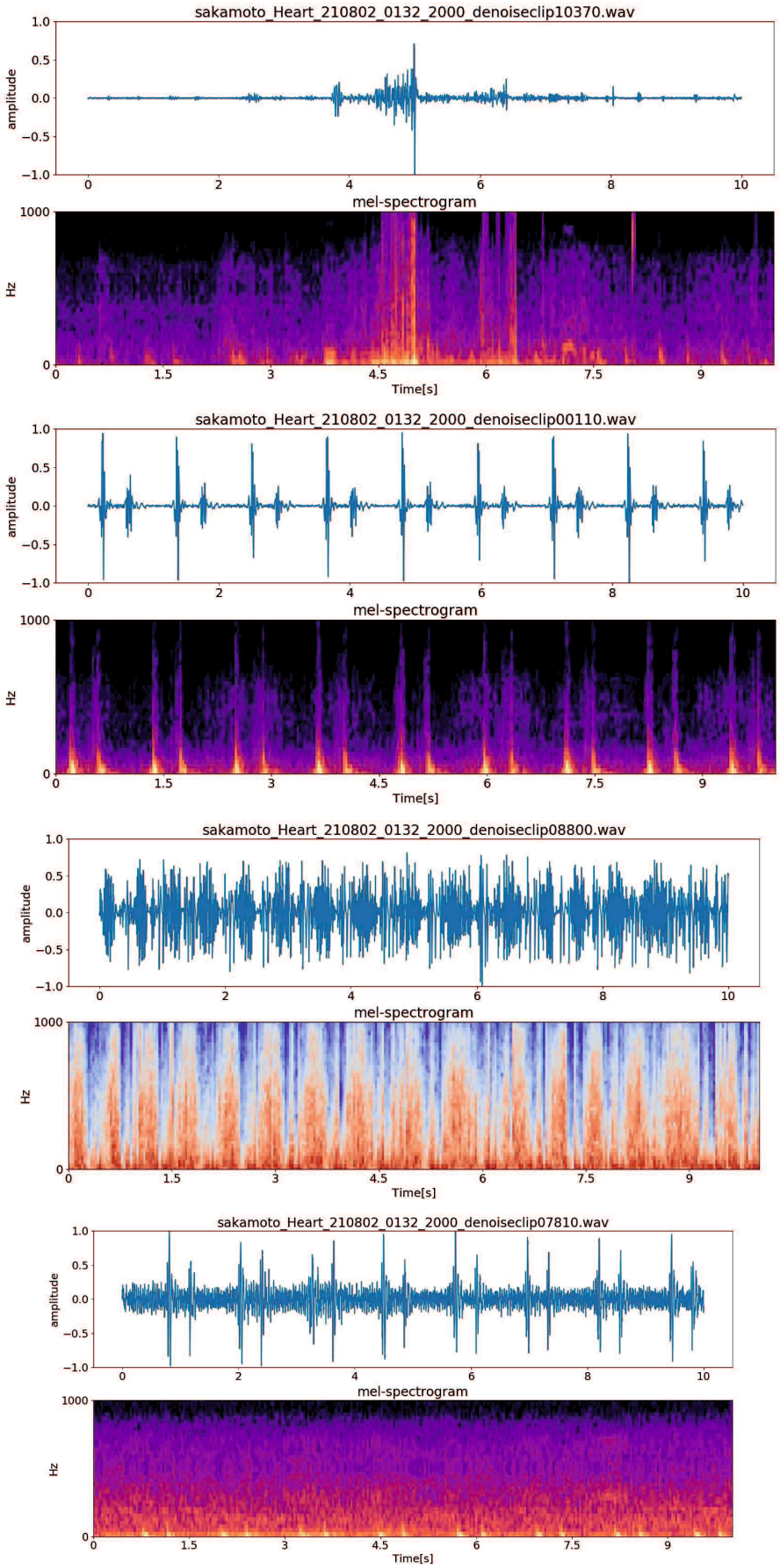

Figure .6.9 representative spectrum of cluster 1,cluster 2, cluster 3 and cluster 4(from top to buttom)

cluster 1: Clips in cluster 1 were heart sounds contaminated by friction noise. A different denoising method is needed to remove noise. This part is analyzable after denoising.

Cluster2: most clips in cluster 2 are high-quality heart sounds. This part is analyzable and can be used for subsequent analysis directly.

cluster 3: most clips in cluster 3 were noise caused by turnover. Heart sound is covered by noise. This part can be removed or discarded.

cluster 4: most clips in cluster 4 were heart sounds mixed with lung sounds. This part's characteristic is that frequency components between 50-2500Hz last for more than 1 second. The heart sound and lung sound separation process needs to proceed, or the lung sound removal method needs to be used to extract heart sound.

6.2 Medical data sharing method based on blockchain technology

6.2.1 Introduction to blockchain technology

The heart sound monitoring system can be used in many communities or laboratories. The system may be deployed on different database platforms, such as Oracle or Microsoft SQL server databases. Therefore the data are scattered in a different database. The data usually need to be collected together for analysis. There is no unified way to access and share them. The data are typically exported from the database and aggregated together by Database Administrators. This is a time-consuming process and is under threat of data leakage. Blockchain is a tamper-proof, anonymous peer-to-peer network where each node has a copy of the full ledger. Blockchain technology can be applied in the health and medical domain to provide a holistic, transparent, whole picture of scattered records [12]. This helps the user to get a full picture of scattered data and sustain crucial trust in the system, meanwhile providing a secure method for users to protect their privacy and share their records[13].

Some research has been done on blockchain technology to utilize its characteristics in the health and medical field to utilize its attributes in the health and medical field. Asaph Azaria built a novel, decentralized system based on blockchain technology to deal with Electronic Medical Records (EMRs). Patients can access their EMRs that spread to various providers and organizations [14]. Qi Xia proposed a blockchain-based system to resolve the problem of medical data sharing among organizations in a trustless environment. All actions done to data were recorded in blockchain in a universal format [15]. Alevtina Dubovitskaya proposed a framework based on blockchain technology for cancer patient care to provide security and privacy-preserve access control over EMR data [16]. Yi Chen designed a storage scheme to securely store EMRs on blockchain and cloud storage, and a service

framework for sharing EMRs was introduced [17]. Hongyu Li proposed a blockchain-based data preservation system to solve the EMRs sharing problem. Users can store and share data with high security on the blockchain framework [18]. Faisal Jamil proposed a novel drug supply chain management system based on Hyperledger Fabric and blockchain technology. This system launched an intelligent contract to manipulate access control to electronic drug records and patient EMRs [19]. In this paper, a new user-oriented blockchain-based data-sharing method among different databases is proposed. Users can access personal records stored in different databases through the semi-private blockchain with their private keys. Heterogeneous records can be processed and retrieved in a unified format. User’s privacy is protected by asymmetric encryption.

6.2.2. Data sharing method

The flowchart of the system is shown in Figure 6.10. Users can record their heart or breathing data with a smartphone or smart bracelet. These data are usually stored in a different database. In this paper, a user-oriented semi-private blockchain is built to link databases together and provide service for users. Each registered user or doctor is one node of the blockchain. Users can use their private key to retrieve all their records stored in a different database with the functionality of blockchain. All records were processed and presented in a universal form.

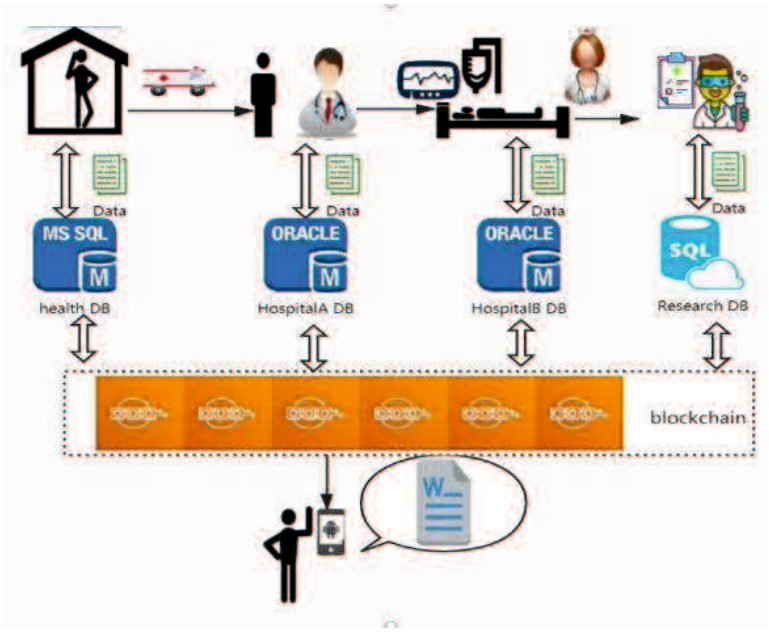

Figure 6.10 Flowchart of the data sharing system

The semi-private blockchain is implemented in this paper to facilitate users' retrieval and improve system security. It has the following advantages of using a semi-private blockchain compared with the public blockchain:

(1) Launching a semi-private blockchain-based application most closely resembles how a company or organization runs a commercial website.

(2) Each node has been certificated, so the risks of being attacked and business failure are low. This simplified implementation and deployment.

(3) The semi-private blockchain uses Practical Byzantine Fault Tolerance(PBFT) algorithm as the consensus protocol. As a result, it consumes fewer resources to reach a consensus state than a public blockchain.

6.2.3 Data Structure

The data stored in a database can be divided into two parts: personal information and medical information. Personal information refers to the part that involves personal privacy, such as name, age, and contact information. Medical information refers to the part other than personal information, mainly including user symptoms, allergy records, heart sound recording files, etc. The data storage of the system is shown in Figure 6.11. Personal information is stored in the blockchain, and medical information is stored in a different database. An index is formed as a pointer to link personal and medical information. The index consists of two parts: the record address and the hash value of the record. The record address points to the access address of the record. The hash value is a verification code of the record. It changes whenever the record is modified. Personal information and index are stored in the semi-private blockchain; medical information is stored in databases that are spread among different databases. These two parts were linked together by the unique index.

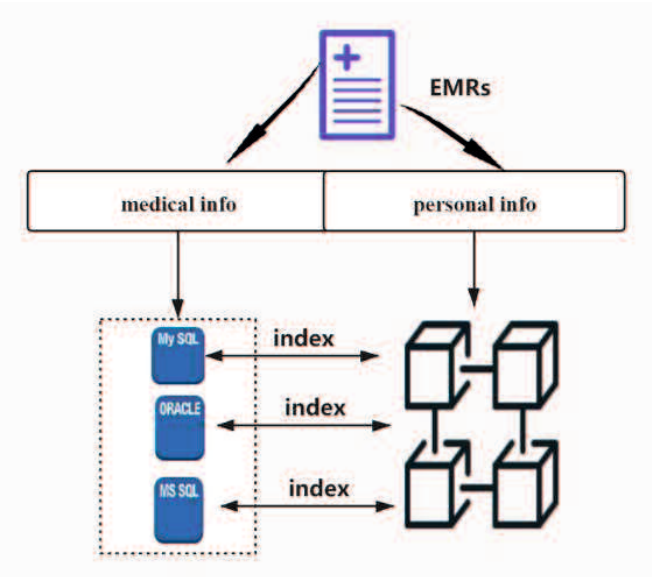

Figure.6.11 Data Storage of System

Each node is equal in the semi-private blockchain, so the data on the blockchain was open to every registered node. While data records contain private information, Asymmetric encryption is introduced to encrypt personal information. Asymmetric encryption provides a pair of keys for each user: the public key and the private key. One can be used as an encryption key, and only the other key can be decryption. The public key is stored in the database and opened for the administrator. The user's private key encrypts medical information, and the administrator can encrypt medical information with the user's public key. The private key is kept secret by the user—the user's vital public decrypts personal information. Only the user himself can use decryption. The hash value of the record will be checked to make sure records were untampered during this process.

6.2.4 Blockchain Structure

A blocklist represents the block's structure in a particular order. The structure of the blocks is shown in Figure 6.12. Two vital data structures used in the blockchain are pointers and linked lists. Pointers are parameters that hold information about the location of another variable. It points to the address of the previous block. Linked lists are a sequence of blocks where each block connects to the last block with the help of the pointer pointing to the former block. The hash algorithm calculated the header address to assure security. The main body of the block is the personal information and record index. There is a bounty part at the bottom of each block.

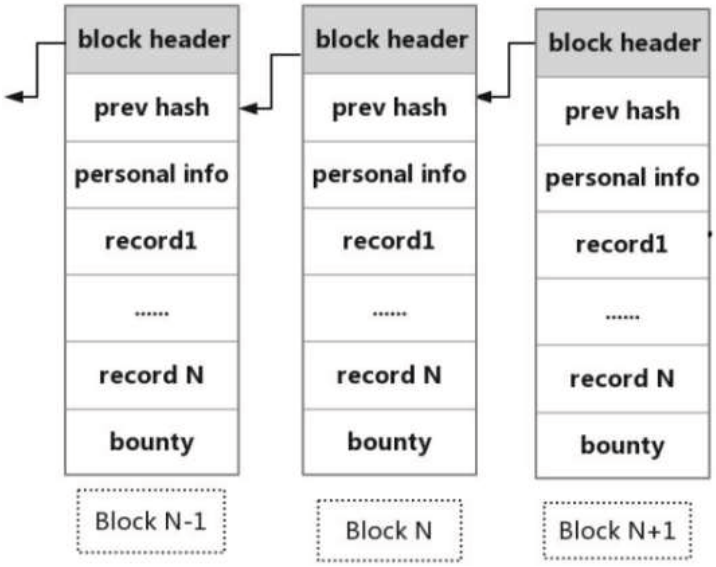

Figure 6.12 Blockchain structure

6.2.5 Retrieving Processing

The user can access all his medical records, which were stored separately through a private key. The retrieval process is shown in Figure 6.13. These records are stored in different databases. A

transform method is introduced to process records in a unified format. Retrieved records are transformed into {key, value} pairs. JSON is introduced as an intermedia. According to the characteristics of a semi-private blockchain, the entire traversal process from the first block to the last block is required to retrieve the information stored in it. According to the user's private key, the traversal process finds the record belonging to the user and decrypts the record with the private key to obtain the index and a hash value of the record. After verifying that the hash value is correct, the records are filtered from the database according to the index. The transform algorithm is used to process records into JSON.

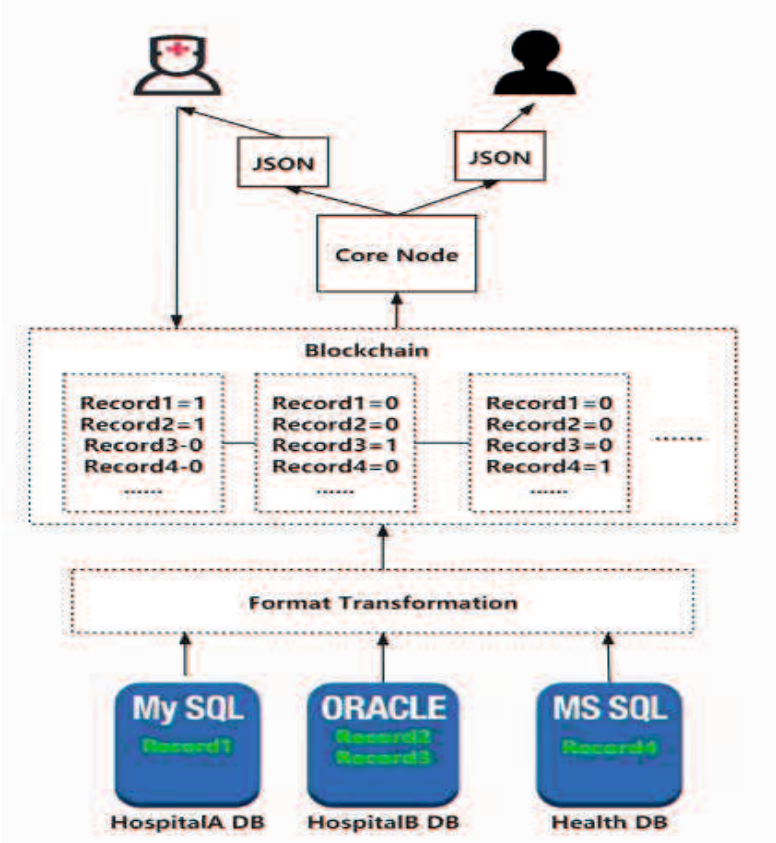

Figure 6.13 Record Retrieve Process

All record fields are changed into five types: empty, string, number, date, and Boolean, based on their original types. The date and timestamp are changed into yyyy/mm/dd format. The number type is changed into float type. The string type is changed into text type. The Boolean type is changed into True/False type. As different databases use different encoding methods, such as Shift_JIS and UFT-8, Unicode is used as a uniform encoding method to make sure all records are stored and shown correctly on the Internet without garbled.

6.2.6.Experiment and Result

Experiments are carried out on Hyperledger Fabric with Python 3.7.3. Hyperledger Fabric is an open-source enterprise-grade distributed Permissioned blockchain technology platform. It supports smart contracts authored in general-purpose programming languages.

Medical and health records are downloaded from the official website of the Department of Health & Human Services(USA) as a test dataset. For the sake of brevity, the records have been simplified. One user (IdentificationID=999-99-9999) is chosen as an example. This user has two records stored in two different databases. The records have a different structure and data format. Record one was created in a community hospital. Record two was created in a bigger hospital after the user was examed in the community hospital. The record stored in the MySQL database is shown in table 6.2. Record two, stored in the PostgreSQL database, is shown in table 6.3. The user uses the private key as a decryption key to decrypt personal and record indexes. Then the indexes are used to get access to records. These two records were retrieved by index from the blockchain. The hash code is checked in this retrieving process to make sure records did not tamper with.

The retrieved result in the JSON file is shown in Figure 6.14. The result consists of three parts. The first part of the JSON file is the user's personal information, the second part is recorded one, and the third part is recorded two. The minus symbol is used to fold sections in the JSON file.

Table 6.2 The Record No.1

Column Name	Value
DoctorID	12000
IdentificationID	999-99-9999
Physical Exam	General Appearance: no acute distress
Medications	HUMULIN INJ 70/30 20 units ac breakfast
Assessment	Sub optimal sugar, control with retinopathy
Timestamp	3/24/2011 12:00:00 AM

Table 6.3 The Record No.2

Column Name	Value
DoctorID	10000
IdentificationID	999-99-9999
Problems	DIABETES MELLITUS (ICD-250.)
Medications	HUMULIN INJ 70/30 20 units ac
Vital Signs	63:130:98.0:72:16:118/60
Orders	Follow-up/Return Visit: 3 months
RecordDate	8/6/2010

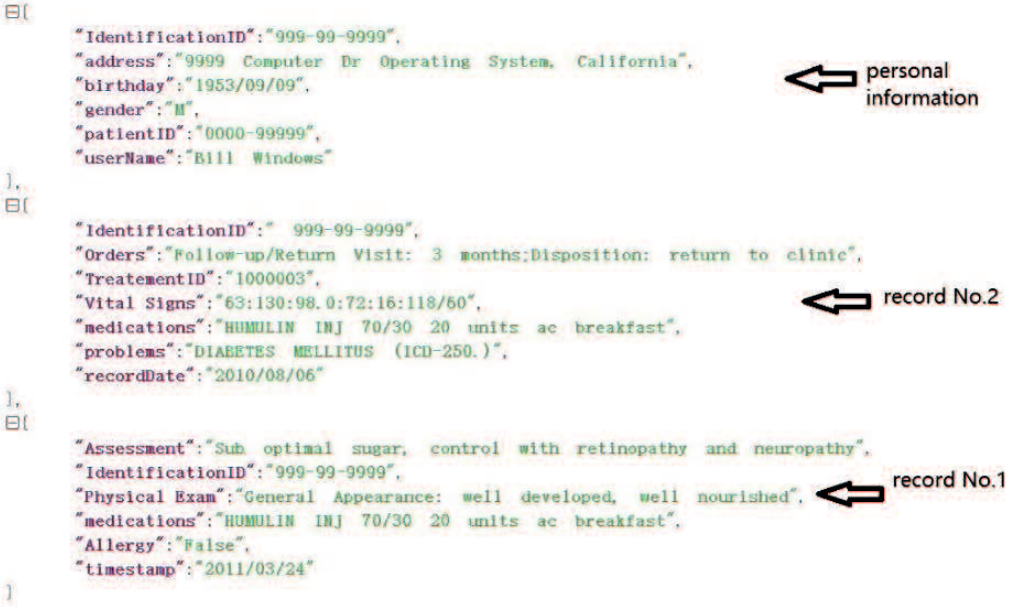

Figure 6.14 Records Retrieved from Different Databases

6.7 Summary

The breathing sound clustering method proposed in chapter 2 is applied to classify heart sounds into clusters and extract analyzable parts for analysis. Record files were segmented into clips, and MFCC was extracted as the feature vector for each clip. AHC was performed on all clips to classify them into clusters and form a dendrogram. The structure of the dendrogram determined the optimal cluster number. Each cluster can be processed using different methods to extract analyzable parts based on its property. Experiments show that the proposed method achieves high accuracy compared with the manually labeled result. However, the cluster number still needs to be determined by humans based on the structure of the dendrogram. Future work needs to focus on the research of cluster number decision criteria that can make the method fully automatic.

A blockchain-based method is proposed to share data stored in different databases. Records stored in various databases are divided into personal and medical information. The record index is abstracted for each record of heterogeneous medical and health information. Personal information and record index are encrypted and stored in blocks to link records together. The user can retrieve records with a private key through blockchain and transform them into a unified form. JSON is introduced as intermedia to process data. The multimedia file, like the medical image-sharing method, will be considered in future research.

References

[1] Aggarwal S, Loomba R S, Arora R R, et al. Associations between sleep duration and prevalence of cardiovascular events. Clinical cardiology, Vol.36,No.11,2013,pp:671-676.

[2] Cao Y, Xu Y H. Effects of sleep disorders on cardiovascular disease. Vol.25, No.1,2020,pp:86-88

[3] Syeda-Mahmood T, Wang F. Shape-based retrieval of heart sounds for disease similarity detection.European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2008, pp:568-581.

[4] Amit G, Gavriely N, Intrator N. Cluster analysis and classification of heart sounds. Biomedical Signal Processing and Control, Vol.4, No.1,pp:26-36.

[5] Ismail K, Salleh S H, Arif A K, et al. Heart Sound Analysis Using MFCC and Time Frequency Distribution[J]. Biomedical Engineering, No.14,pp:946-949.

[6] Park K S, Choi S H. Smart technologies toward sleep monitoring at home. Biomedical engineering letters, Vol.9,No.1,2019,pp.73-85.

[7] Tiwari V. MFCC and its applications in speaker recognition[J]. International journal on emerging technologies, Vol.1No.1, 2010, pp.19-22.

[8] Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol.2,No.1,2012,pp. 86-97.

[9] Le Bel F. Agglomerative Clustering for Audio Classification using Low-level Descriptors, Research Report, 2017

[10] Dave N. Feature extraction methods LPC, PLP and MFCC in speech recognition. International journal for advance research in engineering and technology, Vol.1,No.6,2013,pp.1-4.

[11] Jolliffe I T, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Vol.374,No.2065,2016

[12] Ekblaw A., Azaria A., Halamka J. D. and Lippman A., A Case Study for Blockchain in Healthcare:“MedRec” prototype for electronic health records and medical research data, Proceedings of IEEE open & big data conference, 2016, pp.13-13.

[13] Al Omar, A., Rahman M. S., Basu A. and Kiyomoto S., Medibchain: A blockchain based privacy preserving platform for healthcare data,International conference on security, privacy and anonymity in computation, communication and storage, 2017, pp.534-543.

[14] Azaria A., Ekblaw A., Vieira T. and Lippman A., Medrec: Using blockchain for medical data access and permission management, 2016 2nd International Conference on Open and Big Data IEEE, 2016, pp.25-30.

[15] Xia Q. I., Sifah E. B., Asamoah K. O., Gao J., Du X. and Guizani M., MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. IEEE Access, 2017, pp.5:

14757-14767.

[16] Dubovitskaya A., Xu Z., Ryu S., Schumacher M. and Wang F., Secure and trustable electronic medical records sharing using blockchain,AMIA annual symposium proceedings, 2017, 2017: pp.650.

[17] Chen Y., Ding S., Xu Z., Zheng H.and Yang, S., Blockchain-based medical records secure storage and medical service framework,Journal of medical systems, Vol.43, No.1, 2019, pp.5.

[18] Li H., Zhu L., Shen M., Gao F., Tao X.and Liu S., Blockchain-based data preservation system for medical data,Journal of medical systems, Vol.42, No.8, 2018, pp.141.

[19] Jamil F., Hang L., Kim K.and Kim D., A Novel Medical Blockchain Model for Drug Supply Chain Integrity Management in a Smart Hospital. Electronics, Vol.8, No.5, 2019, pp.505.

Chapter 7

Conclusion

7.1 summary and conclusion

Sleep disorders significantly deteriorate sleep quality and become a hot social issue with the development of an aging society. SRBD is the second one of all sleep disorders. It causes many comorbidities and costs lots of healthcare resources. PSG is the golden standard for evaluating sleep quality and diagnosing SRBD. However, it has many disadvantages, such as the high cost, uncomfortable during sleep, and complicated operation. Other methods for sleep monitoring, such as the RIP or pneumotachograph, their application scenarios are also limited. Therefore, a non-intrusive, easy-to-use, and low-cost sleep breathing monitoring method is indispensable for healthcare in the home environment. The sleeping breath sound is suitable for sleep monitoring in daily life at home. However, sleeping breath sound is not directly related to breathing airflow and is often contaminated with noise. This study aims to develop a breathing sound analysis method for breathing quality evaluation during sleep.

The tracheal sound is preprocessed to remove noise. The TCW and CMW are calculated to segment the breath sound. Based on the envelope, the breathing signal is separated into a low signal part and a normal signal part. The normal signal part is segmented into breathing phases by TCW and CMW. The low signal part is normalized and then segmented. MFCC of each breathing cycle is extracted and classified into normal breathing/abnormal breathing/normal snoring/abnormal snoring/uncertain categories with the AHC algorithm.

As the breath sounds are not directly related to breath airflow, the entire breathing sound file was segmented into 30-second long clips. Seven breathing states are defined and determined based on the classification result and breath regularity. Each clip is classified into apnea, hypopnea, normal breathing, abnormal breathing, normal snoring, abnormal snoring, and event. Therefore the breathing sound is related to breathing states.

AHI is the commonly used parameter to evaluate the severity of apnea. AHI is calculated with different digital signal processing techniques. The apnea index is calculated based on the identification of the breathing states, and the hypopnea index is calculated based on two criteria: the breathing rate and abnormal breathing cycles. The PSG-audio dataset is used for the performance evaluation. Subjects with different OSA severity are selected to evaluate the performance of the proposed method.

Tidal volume is a commonly used parameter to evaluate the breathing ventilation level. However, current methods that can calculate tidal volume need calibration and are only suitable for stable breathing. Therefore, a tidal volume level estimation method using breathing sounds is proposed. Tidal volume is quantitatively calculated for normal breathing and snoring states and qualitatively for apnea/hypopnea and abnormal states.

The proposed classification method is applied for the heart sound analysis. Similar to sleeping breath sound monitoring, the sleeping heart sound is often contaminated with ambient noise or snoring. The whole night's heart sound is classified into several categories based on the recording quality. A block-chain based data sharing method is proposed for the system development.

7.2 Future work

The method's accuracy in this study is affected by other factors, such as ambient noise will cause misjudgment. Also, breathing during sleep is affected by many factors, such as sleep position, pulmonary disease, and body movement. It is challenging to monitor these factors based on breathing sounds. In the future, these factors will be considered to monitor sleep quality. The breathing sound analysis method will be improved with the deep learning methods. More useful parameters will be used to identify the breathing states during sleep. Sleep quality will be evaluated with more factors such as sleep position or head direction. .

Acknowledgement

It is a great honor for me to have Prof. Zhongwei Jiang as my adviser. His profound knowledge, meticulous attitude, and scholarship help me complete research in Japan. Furthermore, he teaches me about my research and allows me to master research methodology. I sincerely appreciate his teaching and advisement.

I thank all the committee professors for their dedication to reviewing my dissertation and valuable comments. I will continue my research with their suggestions.

I want to thank all the teachers and staff at Yamaguchi University. They not only taught me knowledge but also gave me valuable suggestions, which helped me during my life in Japan.

I gratefully acknowledge Prof.Yunlong Wei, Prof. Zhonghong Yan, and all other colleagues of Chongqing University of technology. They encouraged and introduced me to start my research in Japan and gave me valuable advice.

I want to thank all the members of the Micro-mechatronics laboratory, especially Prof. Morita, Prof. Ting Tao, and Dr. Taojin Xu, for their help and encouragement during my study and life in Japan.

I want to thank the Ministry of Education, Culture, Sports, Science, and Technology of Japan, Yamaguchi University, and Tokiwa Engineering Society, which granted me the scholarship to support my life in Japan. These scholarships help me to improve my life, especially during the COVID-19 pandemic.

I want to thank all my friends in Japan. They help me adapt to a new environment and support me when I deal with difficulties. It is a great joy to build a friendship with these friends.

Finally, I would like to thank my family, who supported me and encouraged me to finish my research.

Thank everybody who supported and helped me during my study abroad, as well as express my apology that I can not mention personally one by one.

Wang Lurui