creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학박사 학위논문

수면 호흡음을 이용한 폐쇄성 수면 무호흡 중증도 분류

- Obstructive Sleep Apnea Severity Classification using Sleep Breathing Sounds -

2017년 8월

서울대학교 융합과학기술대학원

융합과학부 디지털정보융합전공

김 재 필

수면 호흡음을 이용한 폐쇄성 수면 무호흡 중증도 분류

- Obstructive Sleep Apnea Severity Classification using Sleep Breathing Sounds -

지도교수 이 교 구

이 논문을 공학박사 학위논문으로 제출함
2017년 4월

서울대학교 융합과학기술대학원
융합과학부 디지털정보융합전공
김 재 필

김재필의 공학박사 학위논문을 인준함
2017년 7월

위 원 장	이 원 종	(인)
부위원장	이 교 구	(인)
위 원	이 중 식	(인)
위 원	곽 노 준	(인)
위 원	이 석 진	(인)

Abstract

Obstructive Sleep Apnea Severity Classification using Sleep Breathing Sounds

Kim Jae Pil
Graduate School of Convergence Science and Technology
Seoul National University

Obstructive sleep apnea (OSA) is a common sleep disorder. The symptom has a high prevalence and increases mortality as a risk factor for hypertension and stroke. Sleep disorders occur during sleep, making it difficult for patients to self-perceive themselves, and the actual diagnosis rate is low. Despite the existence of a standard sleep study called a polysomnography (PSG), it is difficult to diagnose the sleep disorders due to complicated test procedures and high medical cost burdens. Therefore, there is an increasing demand for an effective and rational screening test that can determine whether or not to undergo a PSG. In this thesis, we conducted three studies to classify the snoring sounds and OSA severity using only breathing sounds during sleep without additional biosensors. We first identified the classification possibility of snoring sounds related to sleep disorders using the features based on the cyclostationary analysis. Then, we classified the patients' OSA severity with the features extracted using temporal and cyclostationary analysis from long-term sleep breathing sounds. Finally, the partial sleep sound extraction, and feature learning process using a convolutional neural network (CNN, or ConvNet) were applied to improve the efficiency and performance of previous snoring sound and OSA severity classification tasks. The sleep breathing sound analysis method using a CNN showed superior classification accuracy of more than 80% (average area under curve > 0.8) in multiclass snoring sounds and OSA severity classification tasks. The proposed analysis and classification method is expected to be used as a screening tool for improving the efficiency of PSG in the future customized healthcare service.

Keywords: snoring sound analysis, obstructive sleep apnea, screening test, cyclostationary analysis, feature learning, convolutional neural network, severity classification
Student Number: 2013-31259

Contents

List of Tables

List of Figures

Chapter 1. Introduction

Sleep plays an important role in good health and well-being for individuals. Physical health, brain functions, emotion, daytime performance and safety can be affected by sleep. Sleep apnea is one of the most common sleep disorders and is responsible for a variety of chronic diseases and complications. This thesis describes the studies of sleep disorder related snoring sound and obstructive sleep apnea severity classification using only the breathing sounds during sleep.

1.1 Personal healthcare in sleep

Sleep is one of the life-sustaining activities that make up a significant part of an individual's physical activity. Sleep also plays an important role in improving individual health and quality of life. Sleep is crucial for healthy brain function and emotional well-being. Individual sleep health can directly influence decision-making and problem-solving abilities, emotional and behavioral control [1, 2]. Besides, sleep is a critical factor in physical health, and sleeping health can cause or exacerbate diseases such as heart disease, kidney disease, hypertension, diabetes and stroke [3-5]. Sleep also has a direct impact on individual's daytime performance and safety, and unhealthy sleep can typically lead to a decline in individual productivity and increase drowsiness during driving [6-8]. In the case of the United States, 1,500 deaths per 100,000 car

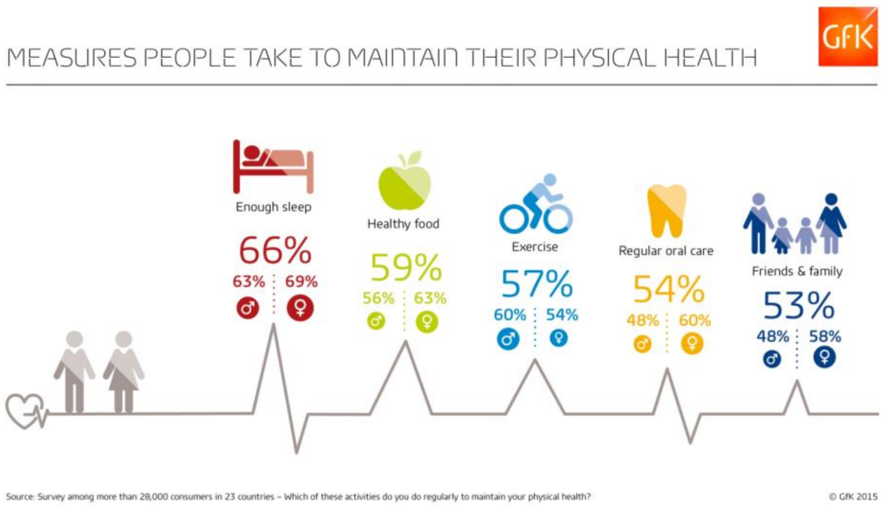

Figure 1.1 Measures people take to maintain their physical health (image source: GfK)

accidents have been reported to be caused by sleepiness during driving[*]. Thus, people are instinctively aware of the importance of sleep health because the sleep health of an individual can be an important factor in the worsening and induction of disease. This is confirmed by the findings of one survey (Figure 1.1) that the most common practice for personal healthcare is sleep[†].

1.1.1. Sleep disorders and diagnosis

In 2005, the International Classification of Sleep Disorders (ICSD) listed 81 sleep disorder symptoms in ICSD-2 with eight major categories [9].

① The insomnias

② The sleep-related breathing disorders

③ The hypersomnias of central origin

④ The circadian rhythm sleep disorders

[*] AAA Foundation for traffic safety (1999), “Why do people have drowsy driving crashes?”
[†] GfK (2015), “Top 5 ways people maintain their physical health”

⑤ The parasomnias

⑥ The sleep-related movement disorders

⑦ Isolated symptoms, apparently normal variants and unresolved issues

⑧ Other sleep disorders

The four main symptoms of sleep disorder in this list are insomnia, sleep apnea, restless legs syndrome (RLS), and narcolepsy* (Figure 1.2). Briefly describing symptoms, insomnia refers to symptoms that are difficult to take or maintain a good night's sleep. Sleep apnea is a symptom of repeated stops or very shallow breathing during sleep. In general, this is sometimes defined as a severe snoring, but it differs from a simple snoring that allows for adequate airflow into the lungs. RLS is a sensorimotor neurological disorder with a very unpleasant sensory symptoms in the legs with an urge to move the legs, leading to chronic sleep disturbances and daytime business disability [10]. The narcolepsy is a nervous system disorder and sleeping disorder in which the sleepiness of the day occurs during the daily life [11]. These symptoms may cause various diseases or complications related to a cardiovascular and nervous system as well as deterioration of the quality of individual sleep.

These different sleep disorders can be diagnosed through the polysomnography (PSG). This is the standard test to find out the cause of the disease in patients with a sleep disorder, measuring and recording various bio-signals such as EEG (electroencephalogram), EOG (electrooculogram), EMG (electromyogram), ECG (electrocardiogram), etc., and finally, polysomnogram is derived.

The PSG test is performed while the patient wears a variety of sensors on the patient's body and sleeps overnight in the hospital. In the test, EEG is used to analyze the sleep stage and evaluate the quality of sleep. Also, we measure

* Cleveland Clinic (2013), “Common Sleep Disorders”

Figure 1.2 Four common sleep disorders

various bio-signals at same time during sleep: muscle tension measurement through EMG, cardiac activity through ECG, respiration through flow sensor, blood oxygen level using a SpO_2 sensor, leg motion and body position change. By using various sensors, it is possible to measure the various physical activity of the patient during sleep, and based on these data, and the doctor can diagnose various sleep disorders objectively.

Recently, a home PSG [12, 13] that can perform a PSG test at home is also being performed in special cases. This relatively new test has the advantages of using simpler sensors and allowing the patient to undergo PSG test in a comfortable and familiar environment. However, it still requires professional assistances for the sensor wearing or attachment.

1.1.2. PSG limitations and customized personal healthcare

PSG is a standard test for symptoms of sleep disorder but contains several limitations. First of all, PSG requires a high cost of up to $ 1,000 for each patient [14]. Besides, as mentioned above, the patient should be examined at night in the hospital wearing various sensors like EEG, EMG, EOG, ECG, etc. This measurement method can cause serious discomfort to the patient. Also, the first night effect (FNE), which occurs due to the patient's failure to adapt to an unfamiliar hospital environment, may result in an inaccurate PSG test [15].

Due to these limitations, attempts to convert traditional PSG test into a more convenient healthcare service can be naturally proposed, and home PSG, which is still in limited use as a professional medical device, can also amplify this expectation. However, there are various limitations in these attempts. EEG, EOG, EMG, and ECG, which are the core measurement components of the PSG, are still professional bio-medical sensors, and it is inevitable to attach them to the human body during measurement. Recently, these sensors are being simplified for use in mobile environments, but it is still difficult to change the fundamental paradigm of measurement. Even if innovative measurement methods are developed and deviate from the problems of sensor attachment or wearing, there is still a possibility of reliability-related difficulties with the sensor itself.

In this situation, a patient screening test for PSG, rather than a direct PSG test, may be a good example from a personal health care service perspective. It is hard for patients themselves to recognize snoring or sleep apnea symptoms during sleep. Usually, the bedroom partner first recognizes the symptom and notifies the patient to check for the disease. However, as the number of single-person households in the world continues to increase*, the probability that individuals will not be aware of their sleep disorder will also increase steadily. Besides, it is also unreasonable and inefficient to conduct a high-cost PSG test on all patients who are aware of certain sleep disorder themselves and come to the hospital [16]. Unlike general illnesses, sleep disorders are difficult for physicians to diagnose patients immediately. In many general cases, a doctor can perform some instant clinical test such as a blood test and can provide quick clinical results to a patient. However, sleep disorders are difficult to diagnose by such rapid examinations, including a medical examination by interview.

In such a situation, PSG screening test would be of great social and economic benefit to both parties. Before a full hospital PSG, this test may be performed in a PSG room of the hospital or a private bedroom environment. In other words, if objective and accurate screening information can be delivered to the medical staff, an unnecessary examination can be prevented in advance through patient selection, so that the reliability of the test can be maximized and the patient's economic loss can be minimized.

Screening does not require complete and accurate diagnostic performance and can provide results using auxiliary information reflecting specific symptoms. In this case, diagnostic biomedical sensors may not be necessary for the screening test. Thus, direct or indirect parameters can be considered to determine a specific sleep disorder and its severity. If software applications and devices for PSG screening test do not require special biometric sensors, we can expect a big expansion in the market for sleep-related customized personal healthcare services.

Finally, questions remain about which of the symptoms of a sleep disorder

---

* Deloitte University Press (2015), "Single-person households: Another look at the changing American family"

should be screened. The requirements for the new customized healthcare services we have considered are as follows:

① Specific symptoms should be highly associated with personal health with a high incidence.

② No additional sensors are required for measurement, and unconscious measurement should be possible.

③ Measurement results and related services will directly help the patient.

④ The function should be able to be practically linked with a specialized institution such as a hospital.

The sleep apnea mentioned in 1.3.1 is the most common sleep disorder. Obstructive sleep apnea is the most common symptom of sleep apnea and can be a direct cause of various chronic diseases and major complications and also can aggravate the symptoms. This symptom produces a variety of sounds, including snoring. Various studies have described the relationship between sounds and the symptoms [17-19]. These sounds can be obtained using various microphones. In the case of a hospital environment, a PSG room usually has a microphone for patient monitoring. Moreover, recently portable smart devices have become common so that sound recording can be performed by an individual himself or herself anytime and anywhere. Therefore, no additional sensors or devices are required for sound recording.

Therefore, screening for obstructive sleep apnea using sound from patients during sleep can directly help many patients who are not aware of the symptoms themselves. If only the sound can be used to examine a specific symptom, the patient does not need any special sensors or devices, so it is economically less burdensome and can be conveniently measured, thereby greatly reducing the

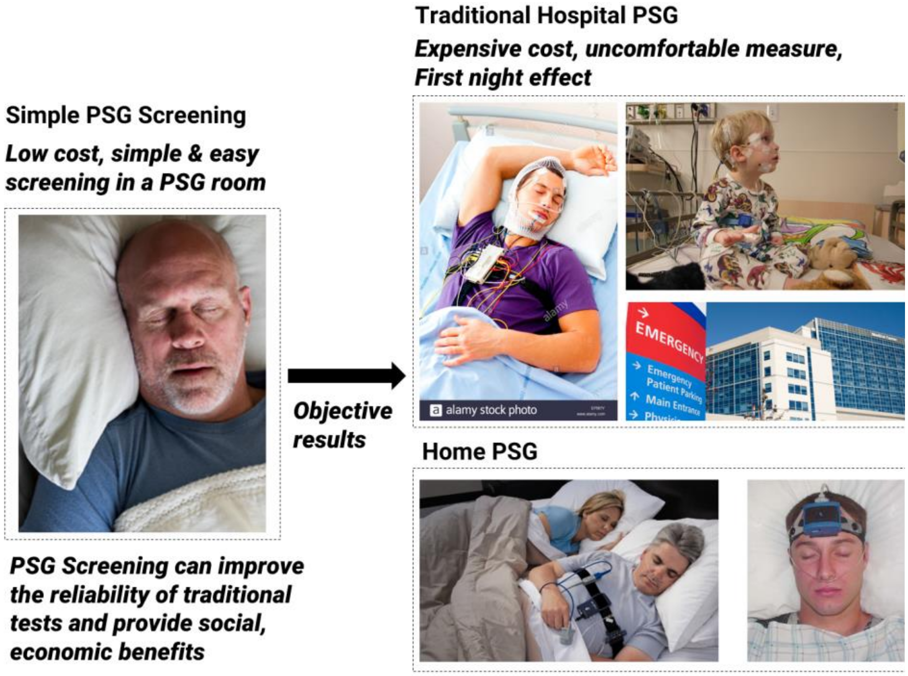

Figure 1.3 Hospital PSG and customized personal healthcare with PSG screening test

patient's resistance to the measurement. The physician in the hospital can perform repeated screening tests on patients and can objectively explain the need for PSG test to the patient based on the results. Consequentially, it is possible to reduce the patient's rejection of the high-cost medical test, and the accurate diagnosis of the PSG can be expected through the selection of the target subject. Judging from these facts, a screening test for obstructive sleep apnea is perfectly corresponded with the new customized personal healthcare service requirements presented above. Besides, sound from patients during sleep can be used as a fundamental parameter for a variety of related healthcare services and can be expected to provide valuable research. These contributions are illustrated in Figure 1.3.

1.2 Existing approaches and limitations

After the first report of obstructive respiration during sleep for the first time in 1965 [20], much research has been done on sleep-disordered breathing. Surveillance of the incidence of obstructive sleep apnea in specific disease groups has been actively conducted in several studies, and the relationship with mortality has also been reported [21, 22]. These studies have reported the incidence of obstructive sleep apnea in a variety of ways and have found this to be a very common disease. This section introduces some representative methods and related studies of a screening test for obstructive sleep apnea.

First of all, sleep survey methods that do not require special measurement devices or data have been proposed. These methods include the Pittsburgh Sleep Quality Index (PSQI) [23] to assess sleep quality, the Sleep Apnea Scale Disorders Questionnaire (SA-SDQ) [24] to evaluate sleep apnea, and the Epworth Sleepiness Scale (ESS) [25] to assess the degree of daytime sleepiness, etc. The PSQI consists of subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, and sleep disturbance. If the total score of PSQI is 5 or more, low sleep quality may be suspected. SA-SDQ scores through 12 items related to age, smoking habits, body mass index, and snoring. The maximum value of SA-SDQ is 60 points, and sleep apnea is suspected when the value is more than 32 points for men and 36 points for women. The ESS assesses the degree of sleepiness in daily life through short questionnaires, with a maximum value of 24 and a score of 10 or more indicating a significant weekly sleepiness. These questionnaires are widely used for individual symptoms but are based on the subjective evaluation of the patient. When a patient should write a questionnaire directly, there is a fundamental limitation of the questionnaire method because a particular patient is impossible or ambiguous to answer a certain questionnaire. Therefore, many related studies have attempted to detect obstructive sleep apnea using various methods based on objective data. Some of the related studies are listed and explained below.

Adrian et al. [26] reviewed a screening test for obstructive sleep apnea using pulse oximetry and noted that it is the most effective and acceptable candidate for low-cost testing than standard Holter monitoring. Based on this, they proposed a sleep apnea screening test method using the clinical score for pulse oximetry tracing pattern, snoring, obesity, and hypertension.

Milton et al. [27] applied a device consisting of a nasal cannula attached to a pressure transducer to the patient's chest via a belt and calculated the apnea-hypopnea index (AHI) value. In this study, they compared the value of AHI in the PSG study with that of the home-based AHI, and confirmed the possibility of screening for sleep apnea.

Martin et al. [28] proposed a sleep apnea screening test using the characteristics of the ECG RR intervals and the area of the QRS complex. The bivariate time varying autoregressive model (TVAM) was used to evaluate the beat-by-beat power spectral densities of the two characteristics. The ECG signals of apnea and non-apnea were classified using K-nearest neighbor (KNN) and neural networks (NN).

Roche et al. [29] proposed a screening test for obstructive sleep apnea using heart rate variability analysis. Various HRV-related variables were extracted in the time domain and the correlation between these variables and obstructive sleep apnea syndrome (OSAS) was confirmed. Multiple logistic regression analysis confirmed that $\Delta[D/N]$ SDNN index (the differences between daytime and nighttime values of mean of the standard deviations of all NN intervals for all consecutive 5-minute segments of the recording) and $\Delta[D/N]$ r-MSSD (the differences between daytime and nighttime values of the square root of the mean of the sum of the squares of differences between adjacent normal RR intervals) were the major predictors of OSAS.

Daniel et al. [30] proposed an accelerometer-based device for a screening test for sleep apnea. A body-fixed-sensor based approach was used, and the accelerometer sensor was immobilized noninvasively on the suprasternal notch of the subject lying in the supine position and the vibration sound was collected from this position. Respiratory, cardiac, and snoring components were extracted from the vibration sound for new sleep apnea diagnosis, and biomedical parameters such as heart rate, heart rate variability, snoring rate, and pitch associated with snores were calculated. These parameters were compared to those obtained with PSG and accurate microphone and also confirmed whether they were suitable portable devices for screening.

Sola-Soler et al. [31] examined the possibility of screening for obstructive sleep apnea using snoring sound. Using the spectrum envelope method, they found a major difference between simply snoring patients and OSA patients. Formant analysis also revealed that SRBD-related patients had higher frequency distribution and found differences in formant frequencies variability between simple snoring patients and OSA patients.

However, most studies have focused on the nature or detection of snoring, and symptomatic studies have mostly focused on determining whether or not patients have obstructive sleep apnea symptoms rather than providing detailed severity information to patients. Most of all, most of the studies were conducted using professional audio recording equipment or using a particular type of microphone. If they use this recording equipment in their experiment, they can get very detailed sound information and apply various analysis methods. However, when the developed algorithm is applied to real service, they can be a high entry barrier.

1.3 Clinical information related to SRBD

Sleep-related breathing disorders (SRBD) are all the symptoms that make it impossible for the lungs to get adequate air during normal human breathing. A typical symptom of SRBD is obstructive sleep apnea (OSA), which includes snoring. This section describes the definitions and features of various SRBDs presented in this thesis. After reviewing the various symptoms associated with SRBD, we will look at a key symptom of this thesis, obstructive sleep apnea.

1.3.1 SRBD

Snoring is included in the SRBD and is typical symptoms that cause discomfort and sleep disturbance to a bedroom partner during sleep. Snoring is a symptom of sleep apnea in a broad sense. Snoring occurs when the air flow of respiration vibrates the tissues behind the throat. Snoring occurs only when sleeping because muscle tension is relaxed and tissues can shiver easily during sleep. It is common in men than in women and varies slightly from study to study, but in general, about 40% of adult men and about 20% of adult females are known to have habitual snoring.

Obstructive sleep apnea (OSA) is a symptom of respiratory pausing over a period due to complete or partial occlusion of the upper airway. Above all, OSA is the most common type of sleep apnea. The upper airway obstruction may be caused by various causes. The most common cause of OSA in adults is changes of anatomical structures and soft tissues in the mouth due to obesity and overweight. When these tissues thicken, they relax in the direction of gravity during sleep to block the airway, which can lead to respiratory pausing. The site of OSA occurrence is known as the soft palate including uvula, the root of

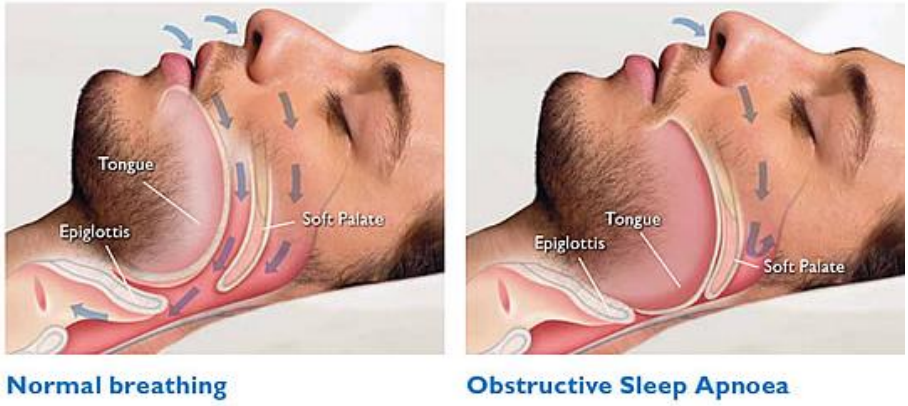

Figure 1.4 Obstructive sleep apnea (image source: Sleep Apnea Institute of Sarasota)

tongue and tonsil. The soft palate containing the uvula is the most vibratory region of the upper airway, and when the patient's uvula and surrounding palate tissue is dropped, the airway is narrowed or blocked easily during sleep. Figure 1.4 illustrates the mechanism of OSA development.

Unlike adults, OSA in children is associated with changes in the size of the tonsils and adenoids, while the direct association with obesity is low. Adenoids are present in the back of the nose, and if it is large, nasal plugging easily occurs. In children, if these symptoms persist, they will continue to have oral breathing. This prevents the growth of upper jaw in anterior-posterior direction, and the nose. In this case, the child will have narrow skeletal structure behind the nose, and this causes more severe snoring or apnea.

The OSA usually has a respiratory pausing for more than 10 seconds, and breathing resumes with a very large snoring sound. To resume breathing, brain’s activity signals are transmitted to various respiratory muscles, which causes brain awakening during sleep. If such a state of awakening continues repeatedly, the quality of the patient's sleep will be greatly reduced regardless of sleeping time.

Hypopnea is a symptom that does not result in complete closure of the upper respiratory tract but has excessive shallow respiration and abnormally low respiration rate. In the case of the hypopnea, the airway is not completely blocked and lasts for more than 10 seconds, which usually involves snoring.

These SRBD-related symptoms reduce the amount of air entering the lungs, thus decreasing the oxygen concentration in the blood, which can lead to various complications.

1.3.2 Obstructive sleep apnea

Patrick et al. [32] described symptoms and common features of sleep apnea during sleep as shown in Table 1.1. Obstructive apnea and hypopnea often occur in the same patient with sleep apnea symptoms. However, regarding treatment, it is meaningless to distinguish between the two symptoms. The PSG test reports the severity of OSA in individual patients using Apnea-Hypopnea Index (AHI) or Respiratory Disturbance Index (RDI) for the number of apneas and hypopneas per sleep hour.

Stenosis or obstruction of the upper airway occurs at one or more sites of velopharynx, oropharynx, and hypopharynx. These sites are fundamentally affected by the elasticity of the nerve roots, the synchronous state of the upper airway muscles, and the sleep stages. It is known that the upper airways muscle decline is common in the rapid-eye-movement (REM) sleep phase, and thus the stenosis and obstruction of the upper airway is prominent during REM sleep [32]. Also, an increase in the adipose tissue of the neck in obese subjects, a tonsil hypertrophy in the case of normal weight patients, or a craniofacial skeletal abnormality may cause stenosis and occlusion of the upper airway during sleep.

Patients with the sleep apnea can be associated with various diseases [33]. Sleep

Distinctive features of syndromes	
Obstructive sleep apnea	Cessation of airflow for >10 seconds despite continuing ventilatory effort
	5 or more episodes per hour of sleep
	Usually associated with a decrease of >4 % in oxyhemoglobin saturation
Obstructive sleep hypopnea	Decrease of 30-50% in airflow for >10 seconds
	15 or more episodes per hour of sleep
	May be associated with a decrease of >4% in oxyhemoglobin saturation
Upper-airway resistance	No significant decrease in airflow (snoring is usual)
	15 or more episodes of arousal per hour of sleep
	No significant decrease in oxyhemoglobin saturation
Feature common to all three syndromes	
Arousal associated with increasing ventilatory effort (as measured with an esophageal balloon)	
Excessive daytime sleepiness	

Table 1.1 Distinctive features of OSA syndromes

apnea patients are known to be at increased risk for diurnal hypertension, nocturnal dysrhythmias, pulmonary hypertension, bilateral ventricular failure, myocardial infarction, and stroke. Sleep apnea also correlates with morbidity and mortality due to cardiovascular and cerebrovascular disease. Repeated increases in sympathetic activity due to sleep apnea can cause hypertension. It is believed that the risk of vascular disease is mediated by interactions between the occurrence of hypoxia, hypercapnia, and influences on an autonomic nervous system of it. Sleep disorders caused by sleep apnea can lead to chronic

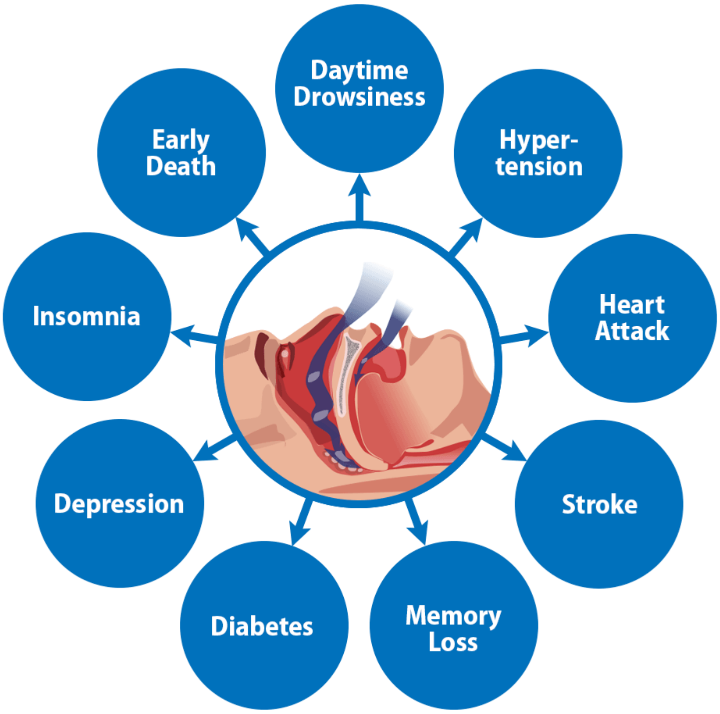

Figure 1.5 Consequences of sleep apnea (image source: nightshifttherapy.com)

sleep deprivation, which can lead to daytime sleepiness, fatigue, cardiovascular complications and cognitive decline. Various diseases and complications associated with OSA are shown in Figure 1.5.

1.4 Study objectives

The purpose of this study is to propose an analysis of SRBD-related snoring and the classification of patients' OSA severity using only sleep breathing sounds for PSG screening test. First of all, the SRBD-related snoring classification experiment was performed using the sleep breathing sounds obtained from the hospital PSG laboratory. Through this study, we have tested whether the proposed algorithm can effectively classify snoring sound events

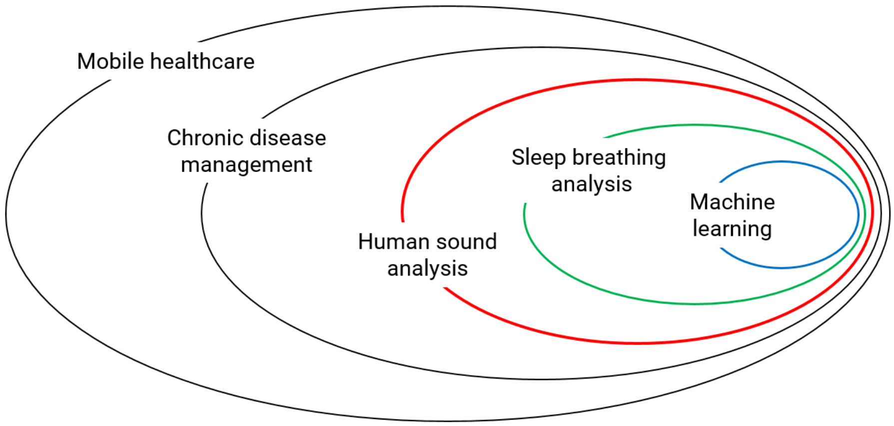

Figure 1.6 Tasks of interest

that are related to SRBD. Besides, we analyzed the OSA severity of individual patients by combining the analytical method mentioned above and the additional time domain characteristics of the patient's overall breathing during sleep. Finally, we tried to improve the performance of the feature learning in the feature representation that was used in the proposed algorithm by using the deep learning technique. To propose an efficient algorithm, we extracted the area of interest from a whole sleep breathing sound and conducted sleep apnea severity evaluation based on it.

1.4.1 Task of interest

Figure 1.6 illustrates the tasks of interest of this study. Our research focused on sleep disorders during chronic disease management in mobile healthcare. We have taken into account the various symptoms that can be detected by actively using sleep breathing sounds of subjects with sleep disorders. Various symptoms of respiratory disorders and obstructive sleep apnea during sleep have been confirmed through various studies. In this thesis, we applied new signal processing and analysis techniques for breathing sounds during sleep, and classify snoring sounds and OSA severity of individual patients using machine learning and recent deep learning techniques. Our primary interest was the ability to distinguish between SRBD and snoring sound events associated with the time domain. We could use this as a basic tool for the patient's OSA severity classification problems if the SRBD-related snoring sound events could be directly distinguished for the sequenced recorded sleep breathing sounds. However, this classification scheme is entirely dependent on the classification performance of the classifier.

If the proposed algorithm shows some possibility to classify SRBD related events, we can use it to generate a particular feature representation for the total sleep breathing sounds. Through this study, we evaluated the severity of OSA in individual patients using only breathing sounds during sleep without special event detector for raw sound. In this way, we can minimize the effect of the snoring event classifier on the final OSA severity discrimination of existing studies, and it will be possible to generate a classification model that includes most of the respiratory sound properties that occur during sleep. To achieve this, it is of utmost importance to produce an efficient and appropriate feature representation of the total sleep breathing sounds.

Recently, the machine learning field based on various data has become more intelligent using the deep learning technology. We have extended the existing traditional hand-crafted feature extraction-based research by applying the feature learning method of deep learning technology. Feature learning was performed by applying various deep learning techniques to the feature representation used in the SRBD event classification study and the OSA severity classification research, and a classification model was created using the derived features. The first study and the second study were conducted in parallel, and we tried to derive the best performance by sharing various information updated through algorithm development in individual tasks. A third study using deep learning techniques confirmed the ability to evaluate OSA severity using sounds from specific regions of interest which were extracted from the total sleep breathing sounds. Besides, we assessed the generalized performance of the classification model so that the study can be used appropriately for actual services.

1.4.2 Contributions

The main contribution of this thesis is to propose OSA severity prediction method for individual patients using only breathing sound during sleep which is recorded in a general recording environment of PSG room. Previous studies have installed the special recording system in the experimental environments and recorded variety of experiments by recording high quality sleep breathing sounds. Many studies have yielded various results through detailed acoustical analysis based on these specially recorded sounds.

For more successful patient-specific healthcare services, patient data should be simply, conveniently, and unconsciously measured from the subject. If the measurement method requires a very special environment and tools, or if it is to be worn on the user's body, the related service is difficult to apply to general users, and expansion of universal service is likely to be limited.

In this study, snoring sounds of patients are classified as symptoms of the respiratory disorder based on sleep breathing sounds recorded with ordinary recording quality without any consideration of sound analysis. We investigate the possibility of classification of hypopnea-related snoring as well as OSA-related snoring, which is the basis of AHI calculation, unlike other previous studies. This has a great significant that it provides a basic tool for directly classifying sleep-related breathing disorder events for calculating AHI in a specific snoring region.

In addition to the classification of these snoring events, we assess the actual patient OSA severity using most respiratory sounds during sleep without detecting snoring events. This is important because it suggests an analytical framework that minimizes the impact of the algorithm's ability to detect the snoring event itself and can utilize more information from most respiratory sounds during sleep. The severity criterion also allows for a detailed prediction of the current state of the patient by utilizing four levels of AHI [34]. This provides more detailed information than the binary classification scheme that distinguishes between simple snoring and OSA-related snoring in existing studies. Therefore, when this method applied to an application for a screening test for PSG, more detailed and various types of screening information can be provided.

One of the main implications of this thesis is the proposal of a new method for analyzing breathing sounds. In this thesis, we applied the cyclostationary analysis method [35], which was used in the field of communication and mechanical engineering, to the audio signal without using the conventional and well-known audio analysis tools. In this study, snoring is assumed to be a signal modulated and released in association with the anatomical region of various upper respiratory tracts [36, 37]. The cyclostationary analysis is a specialized analysis method that can detect the difference between these modulation methods. Therefore, we expected that cyclostationary analysis would have a meaningful effect because we assumed that various modulation characteristics due to a difference in the structure and mechanism of the body organ are released differently when the snoring or breathing sounds related to various respiratory disorders arise from the vocal tract.

Chapter 3: In this chapter, we propose a method to classify recorded sleep breathing sounds from PSG into simple snoring, hypopnea-related snoring, and OSA-related snoring. Based on the various SRBD annotations checked by a medical specialist in the PSG test, snoring sounds are extracted and cyclostationary analysis is applied to each sound. From the cyclic spectrum (CS) function derived from the analysis, we derive the final feature set through dimension reduction, statistical analysis, and feature selection. Based on this, machine learning was applied, and classification experiment was conducted. We propose two different feature extraction methods for CS and examine the possibility of classification of SRBD related snoring sounds.

Chapter 4: In this chapter, we perform cyclostationary analysis and temporal analysis of individual sleeping breathing sounds of individual patients recorded in PSG in each segment without detecting specific events. Through this, a symbolic representation of a person's whole sleep breathing sound is obtained. Temporal analysis derives a stochastic value that transitions from one state to another for changes in the energy pattern of the sleep breathing sound. The cyclostationary analysis obtains the CS mentioned above for each segment of the sleep breathing sounds and applies the dimension reduction, statistical processing, and feature selection method sequentially to derive the final feature set. We concentrated on whether individual snoring can identify the patient's sleep apnea severity based on this representation without focusing on the sound event classification.

Chapter 5: In this chapter, we apply the deep learning method [38] based on the methods mentioned above to enhance the algorithm. In this chapter, we generated the final feature set by applying a very complex process, such as various dimension reduction, statistical analysis, and feature selection techniques, to the CS computed from the segment of snoring or general breathing sounds. However, in this section, we use convolutional neural networks (CNN, or ConvNets) [39] to perform feature learning from CSs for simple snoring, SRBD-related snoring, and other sounds. Then, the region of interest in which a specific energy level transition occurs is extracted from the total sleep breathing sounds, and features are extracted from the audio signal window of each predetermined length using the feature extractor using ConvNets. Finally, based on these features, we perform an experiment to identify the AHI category of each patient using a support vector machine (SVM) [40] and compare the results with the previous methods.

Chapter 2. Overview of Sleep Research using Sleep Breathing Sounds

So far, various sleep studies have been conducted using breathing sounds during sleep. Figure 2.1 shows the general system used to study sleep disorder classification using breathing sound during sleep. In this chapter, we attempt to divide the analytical systems applied in previous studies in the viewpoint of environment setting, analysis and classification methods, and examine the individual contents.

2.1 Previous goals of studies

Various SRBD analysis studies using breathing sounds have been carried out. The patient's breathing sounds were recorded through a microphone installed in the experimental environment and stored in the appropriate file format on the computer system. These stored files were analyzed using various digital signal processing techniques to perform typical tasks such as detection of snoring, classification of simple and OSA-related snoring, and recognition of the anatomical location where snoring occurs. In this chapter, we attempt to divide the analytical systems applied in previous studies in the viewpoint of environment setting, analysis and classification methods, and examine the individual contents.

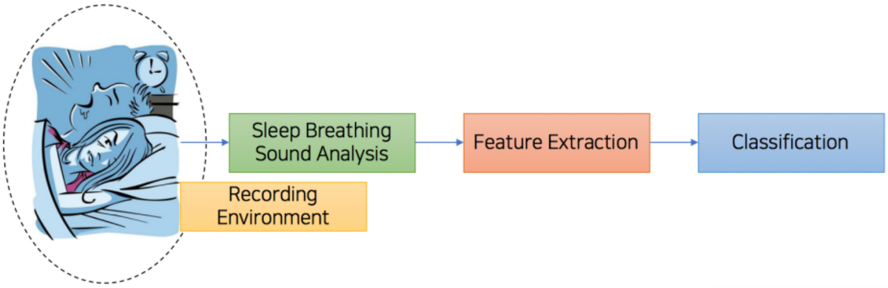

Figure 2.1 Overview of sleep disorder research using breathing sounds

2.2 Recording environments and related configurations

Recording environment is a major factor in SRBD related research using sleep breathing sounds. Clinical studies using sound usually begin by recording the sounds produced by the patient using a microphone. Recorded sounds can be stored in a computer system in various audio encoding formats, and they can be utilized for different purposes.

Sound waves propagating through the air are recorded through a microphone that converts the air pressure change into a voltage. Microphones are typically divided into three categories: condenser microphones, electret microphones, and piezoelectric microphones [41]. Among them, condenser microphones are the most specialized equipment and are known to be the most commonly used equipment in acoustical laboratories. Electret microphones, on the other hand, are used in most small electronic devices and smartphones because they are easy to miniaturize and can be produced at low cost.

There are evident differences between using the expensive professional microphones and using the general-purpose microphones in the viewpoint of building research environments for analyzing sleep breath sounds. Since the former can record high-quality sounds, a very detailed analysis of the sound itself and its consequences can be expected, but the latter may be relatively difficult to study. However, the latter case can be assumed to be similar to the actual recording environments of the general users. With this configuration, we were able to consider various experiments to develop sleep breathing sound analysis solutions that are more appropriate to real life.

In most studies, one or more microphones were placed at very close, about 50 cm, to a patient’s head. Furthermore, there have been some studies that have experimented with various body attached type microphones. In this case, the recordable areas known to date through the studies are the skin of a specific organ, the adjacent parts of the larynx and nasal cavity. Just for reference, so far, there has been no standard for the location of microphones for sleep breathing sounds recording in a related research area.

The physical environment for recording respiratory sounds during sleep is also a major consideration in research. The related literature recommends that microphones should be installed in locations that can minimize the reflected sound of objects such as beds, ceilings, furniture, and walls [42]. In this thesis, we used the sound recorded through the microphone installed in the ceiling of the PSG room as in previous studies. However, unlike previous research, we used a microphone attached to the ceiling surface, about 170 cm from the patient's head. Naturally, this position will minimize the sounds reflected by objects in the room. However, in a typical home environment, it is difficult to install a microphone in the similar location. However, in this thesis, we have identified that breathing sounds can classify the severity of patients with obstructive sleep apnea, even when recorded in a general patient environment, rather than in a professional recording environment, and at relatively remote locations. Based on the results of this research, we determined that the

Figure 2.2 Example of PSG room environment (SNUBH sleep center)

developed system could be extended to the mobile healthcare service environment.

Most of the studies have installed beds and recording equipment in the research laboratory or hospital’s PSG room. To determine the relationship between snoring and sleep disturbances, definite clinical results are needed about the status of patient's sleep disorder. Therefore, it is essential to use sleep system's recorded data and physicians' analysis result during PSG test for developing in-depth analysis method and practical applications. Concerning the recording environment, the PSG room can be regarded as a typical home sound recording environment because it is composed of a common private bedroom with bed, wardrobe, and TV, etc. Figure 2.2 shows the general PSG room environment for this research.

2.3 Sleep breathing sound analysis

2.3.1 Time domain analysis

What the analysis of sleep breathing sounds in the time domain is considered in many studies on the obstructive sleep apnea (OSA). This is because the definition of obstructive sleep apnea is based on the temporal element of complete breathing stop of more than 10 seconds. Therefore, the key to time-domain analysis of audio signals is to focus on the various characteristics of the signal that change over time. A typical example of time domain analysis is an analysis using the crest factor (Equation 2.2) which obtains the root-mean-square (RMS) value of the signal (Equation 2.1) and then divides the largest absolute value of the signal by the RMS value. If the RMS value is high, the signal at this time can be recognized as a peaky-signal [43].

$$V_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} V_i^2} \tag{2.1}$$

$$Crest\ factor = \frac{MAX(|V|)}{V_{RMS}} \tag{2.2}$$

However, these analytical methods may have limitations in deriving OSA-related features and analysis results under certain conditions. To determine the interval of specific sleep breathing event, various breathing or snoring characteristics should be recognized on the time axis. However, these sound events can be easily affected by different noises that occur around the patient during sleep. Although this method has been used in some studies to differentiate sound's properties according to the occurrence location of snoring, it can be considered to be suitable for checking whether a certain target event occurs.

In this thesis, a sound signal is converted into energy values, and these values are quantized into three simple levels of 0, 1 and 2 using several threshold values. Respiratory pausing for more than 10 seconds is a typical feature of the OSA. Therefore, if the zero interval between nonzero levels lasts for more than 10 seconds, the corresponding interval is set as a suspicious section of OSA and set to the additional level value 3. Sequential changes in these level values have been summarized in terms of a representation of the transition matrix and then transformed into probability values. As a result, this feature shows the energy transition trend of a patient's sleep breathing sound over time and it was related to OSA severity [44]. Detailed contents of this feature and analysis method described in Chapter 4.

2.3.2 Frequency domain analysis

Breathing sounds during sleep including snoring are time-varying signals, and the frequency spectrum can be calculated through Fourier transform. The discrete Fourier transform, which is the Fourier transform on the discrete input signal, is defined as follows.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn}, \qquad k \in \mathbb{Z} \tag{2.3}$$

where, N is transform length and k is frequency bin. The frequency spectrum can be calculated for a continuous time and displayed as a spectrogram. Various frequency spectrum analysis methods can be used to identify various information in the breathing sounds.

In many studies using frequency analysis, the fundamental frequency and harmonics of the breathing sound were checked, and the frequency envelopes were analyzed by connecting the peaks in the frequency domain [17, 18, 31, 45].

Wavelet transform was also used for the analysis of respiratory sounds in a related study [46]. Compared with the Fourier analysis method, the wavelet transform is better suited for function expressions with discontinuous sharp peaks and has the advantage of multi-resolution analysis and synthesis.

Mel-scaled spectrograms are often used in speech analysis studies and can be used to reduce the spectrogram dimensions using Mel-scale. Recently, Mel Frequency Cepstral Coefficients (MFCC) [47] is in use as a very suitable tool for speech analysis and classification studies. However, it is known that it does not fully reflect time-varying features for non-stationary non-speech signals [48].

In this thesis, we applied cyclostationary analysis to the data in addition to the existing audio analysis methods. A more detailed description of this is given in next section.

2.3.3 Second-order cyclostationary analysis

The human voice consists of three basic elements: voiced, fricative, and plosive. Vocal sounds are emitted through the area called vocal folds, while the other two are produced at the back of the vocal cords and are unrelated to vocal fold vibration. These two notes are commonly referred to as unvoiced sounds. In the case of fricative sounds, energy is concentrated in a high-frequency band and has characteristics similar to noise [42]. Usually snoring is treated the same as human voice, but snoring is unvoiced, and there are fundamental differences from the voice that it is not produced by the human vocal cords. In other words, snoring is generated through various vibratory activities of the pharyngeal structure, which has been revealed in related research using various endoscopes [49]. Tonsil, tongue root, a vibration of the epiglottis, and flutter of soft palate cause the snoring. Furthermore, since snoring occurs during sleep and the upper airway is not completely autonomous, sound articulation cannot occur. Besides, unlike voice, snoring occurs mainly in the inhalation, so that the sound emission is not the outward but the inward direction [42]. These facts show that snoring sound is much different from normal speech.

Cyclostationary analysis is a widely used signal analysis method in the field of communications and mechanical engineering. In a broad sense, if the autocorrelation of a signal is a periodic function, the signal is called cyclostationary. A modulated signal has a cyclostationary characteristic. In the field of communication, the periodicity embedded in a signal frequently occurs in training sequences and cyclic prefixes. Such periodic characteristics included in a transmission signal can be utilized in a receiver [50]. It is calculated using autocorrelation function and power spectrum density. The actual implementation is to compute the correlation with itself (autocorrelation) using the power spectral density obtained from the received signal. The derived function has one cyclic spectrum with a cycle frequency. Based on this theory, many studies extracted various features by examining the distribution of peaks or calculating the frequency of peaks in their experiments. In a broad sense, noise or interferences are treated as stationary signals and have no spectral correlation characteristics so that this method can be used as a suitable characteristic detector in low SNR environments and when there is a change in noise power.

So far, there has not been much use of cyclostationary analysis in the healthcare field. As a representative related study, Aya et al. [51] performed cardiac arrhythmia classification using cyclostationary analysis method and found that this method showed high classification performance for specific arrhythmia types. Although not many examples were actively used in the sound field, Phani et al. [48] conducted an experiment to compare the performance of the cyclostationary analysis and MFCC for the environmental sound classification system. Sound processing is mainly focused on voiced sound, which includes human voice, voice processing, and music analysis. In this field, MFCC is one of the most popular and widely used analytical methods. However, aforementioned environmental sound recognition task that included the bird sound recognition problem did not produce better results than the cyclostationary analysis method.

In this thesis, core features are extracted based on the cyclostationary analysis of sleep breathing sounds. The SRBD-related breath sounds such as snoring were assumed to be produced by the sound modulation of the anatomical factors such as tonsil, tongue root, and palate. Therefore, the cyclostationary analysis is expected to be an appropriate tool for analysis of sleep breathing sounds.

2.3.4 Features from sleep breathing sounds

Although developing new analysis method of sleep breathing sound is important, from the results of these methods, extracting characteristic features from SRBD related sounds and performing actual target task can also be one of the major processes in the breathing sound classification or OSA severity classification system. To date, many studies have extracted various features from respiration sounds through different analysis methods and classified them according to specific purposes. In this chapter, we examine different methods of extracting features using acoustical analysis methods for sleep breathing sounds in different clinical trials. Finally, we also examine feature learning which is getting much attention in the field of machine learning recently.

2.3.4.1 Discovery of features based on clinical tests

Respiratory sound analysis through clinical tests can be summarized using results from sleep nasal endoscopy. Most studies suggested acoustic features according to the anatomical characteristics and locations of the snoring occurrence. These studies expected that the snoring improvement or change of snoring according to the operation related to snoring treatment could be judged through sleep nasal endoscopy and breathing sound analysis. However, other related studies have raised fundamental questions about the methodology through sleep nasal endoscopy, because they found differences between the features of breathing sound analysis of natural and sleep-induced snoring.

Quinn et al. [52] found in their studies that there is a difference in waveform and frequency patterns between palatal flutter and tongue-based snoring. The result shows that the location of snoring occurrence makes the differences between analyzed features. Hill et al. [37] conducted a study of snoring using crest factor as mentioned above. They argued in this study that areas of severe snoring occurrence that were detected during a one-night nasal endoscopy could not be representative. They also analyzed the snoring sounds occurring at different times on the same day's test. They found that, in some patients, the mechanism of snoring generation may be altered, regardless of changes in snoring occurrence sites. Agrawal et al. [53] compared snoring in natural and drug-induced sleep based on the fact that midazolam or propofol is used to induce sleep when patients undergo a sleep nasal endoscopy for snore tests. They found that sleep nasal endoscopy does not fully reflect the nature of snoring by identifying differences between the two snoring in their experiment. Jones et al. [54] also found that measured loudness of the snoring sound increases with increasing levels of sleep induction, revealing a clear difference between natural snoring and sleep-inducing snoring.

2.3.4.2 Feature extraction through spectral analysis

In the studies up to now, many characteristics of sleep breathing sounds have been extracted through spectral analysis. As representative examples, mean power ratio (MPR), center frequency (CF) and peak frequency (PF) of the spectrum were used to distinguish characteristics according to the location of snoring occurrence in the upper airway.

Schafer et al. [55] found that in the case of simple snoring, the frequency spectrum has low-frequency components and a large number of harmonics. Saunders et al. [56] divided snoring into the palatal-based snoring and tongue-based snoring based on the range of CF values, and they also found that this could be a substitute for the sleep nasal endoscopy, depending on the results of acoustical analysis of snoring. Agrawal et al. [53] extracted spectrum-related information, MPR, CF, and PF from 12 palatal-related snoring patients and studied the characteristics of ordinary snoring and sleep-induced snoring. Jones et al. [54] analyzed the features of sleep-induced and natural snoring through energy ratios for low-frequency subbands. Herzog et al. [57] identified simple and OSA-related snoring using peak intensities from the power spectrum for inspiratory snoring and found that simple snoring contains multiple harmonic intensity peaks, and has peak intensities in the 100-300 Hz band. Besides, they found that peak intensities existed above 1,000 Hz for OSA-related snoring. Ng et al. [58] thought that general power spectrum analysis is not sufficient for SRBD analysis because snoring signal has non-Gaussian and nonlinear behavior characteristics. Accordingly, they attempted to detect OSA through nonlinear properties using bispectral analysis. Bertino et al. [59] looked at the difference between the two symptoms through formant analysis based on the fact that snoring and sleep apnea related surgical procedures alter the anatomical structure of the upper airway and the resonance characteristics of the vocal tract.

2.3.4.3 Feature learning via deep learning techniques

With the recent development of deep learning related research, there have been many attempts to learn the feature through neural networks without hand-crafted manner in various fields. In particular, various attempts have been made in the area of image processing and recognition, and recently, related studies have been expanded to the area of sound processing.

Honglak et al. [60] proposed a deep learning approach that can be applied to a wide range of audio recognition tasks. After converting audible data such as unlabeled speech and music into a spectrogram, they performed feature learning using convolutional deep belief networks (ConvNets), and obtained the learned feature representation. The feature representation derived from this study is claimed superior to the widely used spectrogram and MFCC for various existing audio classification tasks. Also, these learned features were expected to be utilized in various audio recognition tasks since they showed higher performance than other features even when the number of training examples with labels was small.

In this thesis, we converted the preceding cyclic spectrum (CS) derived from the cyclostationary analysis of a snoring sound into an image form [44, 61]. We applied ConvNets to these images and learned features from CS. The classification results obtained from these learned features were compared and evaluated regarding classification performance with results of previous studies which extracted a feature from CS in a hand-crafted way.

2.4 Sleep breathing sound classification

In various research, data sets consisting of sleep breathing sounds were classified according to different purposes using the different data analysis and feature extraction methods. Tasks can be broadly divided into two classification tasks: snoring recognition among the various respiratory sounds, detecting specific pathological symptoms such as an OSA. In previous studies, OSA event detection and classification are necessary processes to assess whether a particular patient has OSA symptoms, and furthermore, to evaluate the OSA severity of an individual patient.

As described above, the OSA diagnosis of individual patients is evaluated based on the AHI value derived from the hospital's PSG test. Most studies have simply assessed whether a patient is an OSA patient by analyzing sleep breathing sounds [45, 62, 63]. At this time, the AHI threshold which is the criterion for dividing the patient into two classes (OSA/ non-OSA) was arbitrarily selected to adjust the sensitivity and specificity of the results. However, some studies have classified OSA severity using sleep breathing sounds according to four-level criteria of AHI [64].

In this thesis, we classify the patients based on the multiple indexes corresponding to the standard AHI severity category through the proposed sleep breathing sound analysis method. More detailed patient prediction of OSA can provide more detailed information in real-world applications than other methods. Moreover, the analysis method using only sound can make the measurement manner easy and simple, so repeated measurements allow to analyze the trends related to OSA.

2.5 Current limitations

The purpose of our study is to investigate the possibility of a screening test for PSG test through the analysis of sleep breathing sounds. This screening test should be able to be performed simply and repeatedly when we consider future mobile healthcare-related services. However, to date, many studies have performed the sleep breathing sound recording process by attaching a special microphone to the body [17, 19, 46, 53, 62-65] or using a high-performance microphone placed in close distance to the subject’s head [18, 19, 37, 55-58]. Recorded sleep breathing sounds in this way have a high SNR, and therefore, the minute characteristics included in the sounds can be analyzed in detail. However, professional recording environment or requiring special device can make it difficult to carry out PSG screening test in ordinary environments. We believe that PSG screening test should ultimately be applied to mobile healthcare service for self-awareness and treatment of sleep disorder and also be realized via the personal mobile devices. Unfortunately, we have not been able to conduct experiments based on typical home bedroom environment and data recorded from mobile devices. However, we performed experiments in a recording environment that can offset the influence of the position and distance of the microphone somewhat in a physical environment that is not specifically controlled than any previous experiment. The PSG room is similar to a typical bedroom environment, but sleep technicians who administrate PSG tests may visit several times during sleep to adjust the sensor or resolve complaints of the patient. To remove the speech or noise caused by the visit of the technicians or the patient behavior, in some case of experiments, the breathing sounds which were included in the particular sleep stages were used. Therefore, to apply the specific algorithm to real healthcare services in the future, it may be necessary to analyze the sleep stage. However, since we chose breathing sounds mainly at deep sleep stages, if needed, sleep stage filtering method can be easily implemented by using body movement information from a built-in accelerometer in mobile healthcare devices.

To summarize, the experimental limits of this thesis are as follows:

① Various types of microphones and various installation sites have not been considered for sleep breathing sound recording. The microphone used in the experiment is installed on the ceiling of the PSG room, and the distance from the patient's head is about 1.7 meter.

② Data analysis did not consider real-time processing. The audio data used in the experiment is the audio files extracted from the full recorded video for room monitoring during PSG.

③ The breathing sounds recorded at specific deep sleep stages were used in certain experiments to remove speech or various unintended noises.

Based on above limitations, the goals of this thesis can be clearly defined.

① Compared with the previous experiments, we investigate the possibility of SRBD-related snoring and OSA severity classification using the sleep breathing sounds which are recorded with the most common recording environment and equipment in the PSG room.

② Ensures task performance that enables PSG screening from recorded sound data in a relatively uncontrolled environment. Based on this, it will be available that predicting the possibility that developed analysis algorithms and its frameworks can be easily applied in other experiment environments.

In the following chapters, we will describe the details of the three different studies that we have done and their results.

Chapter 3. Multiple SRDB-related Snoring Sound Classification

In this thesis, we perform two kinds of sleep breathing sounds analysis tasks. The first is to distinguish SRBD-related snoring events from snoring sound units extracted from full-night sleep breathing sound, and the second is to obtain a specific feature representation from long-term sleep breathing sound and use it to classify OSA severity of individual patients. This chapter describes the first task.

3.1 Introduction

Sleep-related breathing disorder (SRBD) which shows complete or partial upper airway obstruction including snoring, is very common symptoms that lead to person's sleep disturbance and bed partner's discomfort. In some severe cases, it causes or worsens the patient's excessive daytime sleepiness and various cardiovascular and neurovascular complications [33, 66].

In this chapter, we have performed experiments to classify three kinds of snoring sounds. A specific length of OSA related snoring sounds was extracted from the patient's sleep breathing sounds and verified that the individual sounds could be classified based on a particular feature representation.

Many studies have been conducted to detect snoring events from sleep breathing sounds [57, 64, 67]. The detected snoring sounds were classified into simple (non-OSA) snoring sound or OSA-related snoring through various methods. Most existing studies detect the occurrence of snoring preferentially, and then set the snoring interval according to established criteria. Finally, they identified snoring type of the interval with various analysis methods which described in Chapter 2. On the other hand, our study investigated whether the chunks of signal per unit of a window could be classified as a snoring type representing various symptoms of sleep disorder. In other words, we focused on developing a feature extractor that accounts for a given sound section rather than developing a snoring sound detector.

In this chapter, we have conducted extended tasks that classify three types of snoring, rather than existing research: simple snoring, OSA-related snoring, hypopnea related snoring. Creating a classification model that can distinguish the three snoring types has many implications. If a model with perfect classification performance can be generated, this can be the major tool for calculating AHI from a patient's overnight sleep breathing sound. However, creating a complete classification model is a very difficult or impossible task. Thus, a model or a feature of the previous task that can lead to achieving reasonable performance can help generate another new feature representation that reflects the characteristic of the patient's overnight sleep breathing sounds. Furthermore, by performing the window-based SRBD related snoring classification task, a single feature representation of the breathing sounds contained in individual windows can be generated. This means that when the task is extended to overnight breathing sounds analysis task, feature representations of all individual sound intervals can be collected. More details on this can be found in Chapter 4 of this thesis.

We used the cyclostationary analysis described in Chapter 2 to extract features that can distinguish the three types SRBD related snoring. Cyclostationary

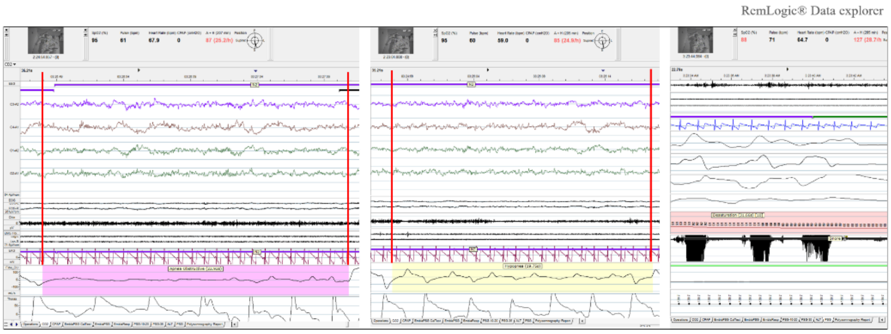

Figure 3.1 Examples of the PSG Software's annotation (RemLogic® Data explorer)

analysis changed the snoring sound to the specifically transformed matrix and based on this, and two feature representations were generated. Then, we also made various classification models using various classifiers and compared their performance.

We start in Section 3 with a system architecture, which includes data specification, analysis method, feature extraction and classification method. In Section 3, we describe the evaluation process of a classification model. Next, in Section 4 and 5, we describe the result and conclusion of this experiments.

3.2 System architecture

3.2.1 SRBD-related snoring event database and preprocessing

The breathing sounds for this study were recorded during hospital PSG tests which were conducted in the sleep laboratories at the Seoul National University Bundang Hospital. The number of involved patients was 12 and the average recording time was 7 hours. Each examination process in a sleep laboratory was recorded using a microphone with a frequency band between 20 and 2000 Hz (SUPR-102, ShenZhen YIANDA electronics) placed on the ceiling above a

Type	Extraction criteria	Duration
Simple snoring	Annotated section's duration	2 seconds
Hypopnea-related snoring	Annotated section's duration or no annotation but indubitably, snoring	2 seconds
OSA-related snoring	snoring sounds that occurred immediately after the OSA annotations	2 seconds

Table 3.1 Criteria for three types of snoring sound extraction

patient's bed at a distance of 1.7 m. We extracted sounds from the recorder through a widely used multimedia library, FFmpeg [68]. Then, these sound files were resampled to 8 kHz for effective analyses and filtered by a spectral subtraction algorithm [69] to reduce unintended noises. Analyses of the PSG data were performed using a sleep diagnostic software, RemLogic® (Natus Medical Inc., USA). The various clinical events were eventually analyzed by medical professionals and the point of occurrence and duration of the related events were displayed and stored with the form of annotation in the PSG software. In this experiment, we considered three types of snoring: simple snoring, hypopnea-related snoring, OSA-related snoring. Using these annotations, three types of snoring events, simple were extracted from the sound data extracted from monitoring video which is time-synchronized. Based on these extracted events, 50 snoring sounds were randomly selected for each type, and the duration of each sound was 2 seconds.

There were a few considerations in annotation-based snoring sound extraction. The OSA is an event in which the breathing is stopped for a few seconds or a few minutes. Thus, related annotations written by the experts span the silent section on the synchronized PSG sound signal. For that reason, we collected snoring sounds that occurred immediately after the silent apnea period, not the

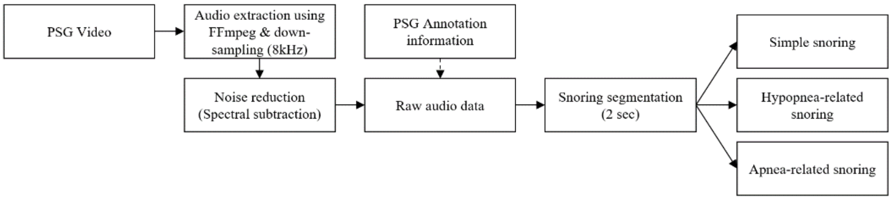

Figure 3.2 Preprocessing of this experiment

sounds of the annotated OSA section. On the other hand, hypopnea and simple snoring related sounds were directly extracted based on the time duration of its annotated sections. The following Figure 3.1 shows an actual example where each snoring event is displayed on the real PSG software.

In the individual SRBD related snoring segments extracted by the above process, we manually selected 2 seconds duration snoring sound. As a result, we could get an experimental data set which consisted of total 150 individual snoring sound files with three different classes. The Table 3.1 shows the criteria for three types of snoring sound extraction methods and the preprocessing of this experiment can be summarized with the diagram of Figure 3.2.

3.2.2 Feature extraction method

From extracted SRBD related three types of snoring sounds, we attempted feature extraction through the cyclostationary analysis method. In Chapter 2, we briefly described this analysis method. If the section of the thesis in which the cyclostationary analysis applied appears, we will individually explain in detail how it has been applied to each experiment. In this study, we transformed the snoring sound of 2 seconds length into cyclic spectrum using cyclostationary analysis method and performed feature extraction through principal component analysis or integral image conversion of a covariance matrix.

3.2.2.1 Spectral Correlation for the Snoring Sound

In the signal processing research area, most studies considered a given signal to be stationary. However, most of the signals that occur in nature are non-stationary [70]. The extracted snoring sounds in this study were considered non-stationary data including repetitive and complex waveforms. The sounds were analyzed offline and we tried to extract the cyclostationary properties from theses sounds. In general, one signal is cyclostationary when the signal is non-stationary, and its statistical characteristics are varying periodically in time domain [71]. When we conduct a cyclostationary analysis, if a signal x(t) can decompose to several sine wave components through a non-linear transformation of order n, x(t) is defined as nth order cyclostationary process. The non-linear transformation can represent a correlation function including time lag of x(t) using a quadratic transformation. Therefore, conducting a cyclostationary analysis can show the hidden periodicity of the signal of interest and can extract the hidden features of data. On the other hand, typical noises do not indicate any cyclostationary properties. Therefore, this analysis method can be considered to be robust regarding noises. According to the above definition, a non-linear transformation of signal x(t) generates a non-zero spectral line when the α is greater than zero. Alpha means cyclostationary frequencies which can be obtained by $n/T_0(n \in \mathbb{Z})$. Combining the above concepts allows inducing cyclic autocorrelation function (CAF) R_{xx}^{α} which is represented in equation 3.1.

$$R_{xx}^{\alpha}(\tau) = \lim_{T \to \infty} 1/T \int_{-T/2}^{T/2} R_x(t,\tau) e^{-j2\pi\alpha t} dt \quad (3.1)$$

where, $R_{xx}(t,\tau) = E\left\{ x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) \right\}, \ \alpha = \mathrm{n}/T_0 \ (n \in \mathbb{Z})$

The CAF can be considered to be a measurement of the correlation between the frequency shifted versions of x(t). The CAF can also be a periodic function which represents the second-order periodicity of the x(t) and assumes a cyclostationary process. As we obtain the necessary information of periodic function through Fourier transform, we can acquire the cyclic spectrum (CS) by performing the Fourier transformation of the CAF. The CS is represented in equation 3.2.

$$S_{xx}^{\alpha}(f) = \int_{-\infty}^{+\infty} R_{xx}^{\alpha}(\tau) e^{-j2\pi f\tau} \tag{3.2}$$

We also can represent the discrete version of CAF and CS of signal x[n] which has fixed time lag l. They can show as equation 3.3 and 3.4.

$$R_{xx}^{\alpha}(l) = \lim_{n \to \infty} \frac{1}{N} \sum_{m=0}^{N-1} x[m] x^{*}[m+l] e^{-j2\pi\alpha m\Delta m} \tag{3.3}$$

where N is the number of samples of signal x[m], Δm is the sampling interval.

$$S_{xx}^{\alpha}(f) = \sum_{l=-\infty}^{\infty} R_{xx}^{\alpha}(l) e^{-j2\pi f l \Delta l} \tag{3.4}$$

As mentioned in the above, if the spectral components greater than zero exist in signals of interest, the signals can be considered to have a second-order periodicity, which also can be regarded as a hidden periodicity. The hidden periodicity was comprised of many spectral correlation coefficients. Therefore, we selected these coefficients for generating features of target signals. Figure 3.3 shows results of the CS of representative signals corresponding to simple snoring, hypopnea-related snoring, and apnea-related snoring. Although results were not completely same patterns in individual three kinds of snoring, they showed that spectrum components in cycle frequency domain of each snoring might have different cyclostationary properties.

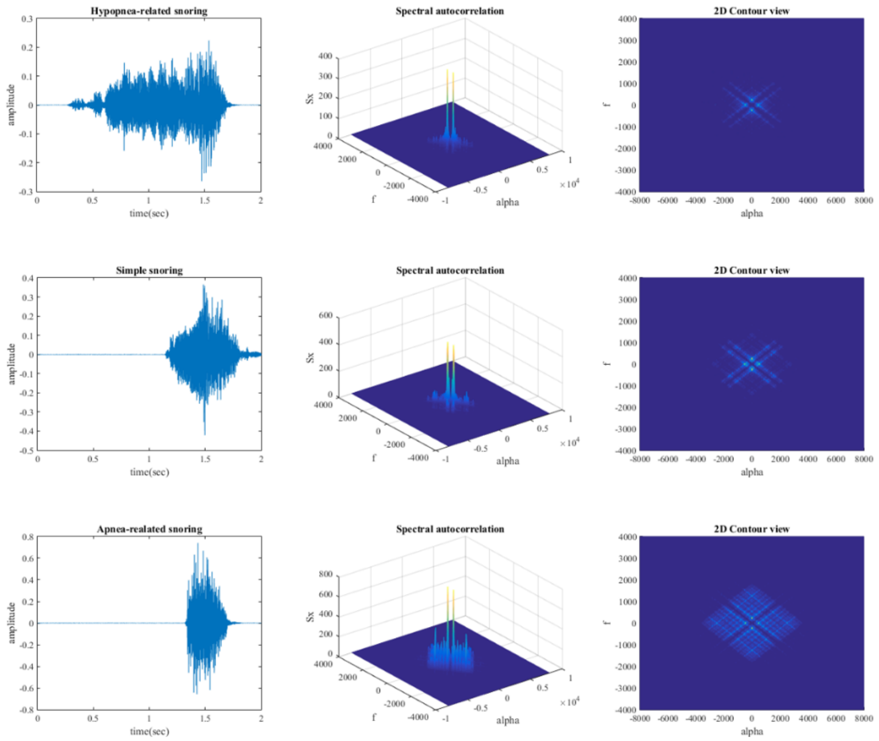

Figure 3.3 Comparison of the cyclostationary component in 3 types of snoring (subject 1)

3.2.2.2 Feature extraction from cyclic spectrum

These cyclostationary features were extracted from foregoing three symptoms-related snoring sounds with two second signal window. First of all, CS was calculated from each snoring sound, and this built a spectral correlation matrix (SCM). The SCM was a result of the CS which includes magnitudes corresponding to frequency and cycle frequency domains. The region of interest of the SCM which was partially removed the zero magnitude regions was extracted using the method of the previous related study [51]. Based on this filtered SCM, we calculated the max, mean, standard deviation of each cycle frequency (column) according to whole frequency range (raw). The SCM was generated for each frame and results were averaged by the total frame numbers. Then, a feature set was produced by attaching each of the statistical value of the coefficients. The dimension of the feature set showed definitely high dimensionality (35,901). Therefore, we conducted principal component analysis (PCA) and obtained a new dimension-reduced version of feature set which includes 149 PCA coefficients. A pseudo-code for a feature extraction so far is given in Algorithm 3.1. In the code, cyclic resolution (dalpha) is calculated by 1/T, where T is the observation time of the data, and spectral resolution (df) is calculated by 1/Tw, where Tw is the small window time (sliding FFT time length) along the entire observation time interval dalpha [72].

Algorithm 3.1 A Feature Extraction Method 1 Pseudo-Code
input A snoring sound S, frame rate Fr, frame duration D, average ROI coordinate set O_C, window scaler V
1: Calculate frame size Fsize: Fsize = Fr×D
2: Extract frame data, FD by Fsize: FD ← S[$Fsize$]
3: Cyclic resolution dalpha = $1/(\frac{\text{Fsize}}{\text{Fr}})$, Spectral resolution df = $1/(\frac{\text{Fsize/V}}{\text{Fr}})$
4: Calculate the cyclic spectrum (CS): magnitude (S_x) = CS(FD, df, $dalpha$)
5: Crop the magnitude for extraction the region of interest $S_x crop = S_x[O_C\{x_1, x_2, y_1, y_2\}]$
6: Calculate averaged maximum, standard deviation, mean of $S_x crop$ $avgS_x crop = [max(S_x crop), std(S_x crop), mean(S_x crop)]$
6: Principal Component Analysis: $avgS_{x_crop}$score ← PCA($avgS_x crop$)
output $avgS_{x_crop}$score

Another approach for feature extraction of cyclic autocorrelation function has been attempted. Several studies have used covariance matrices as object descriptions of an image. Assuming that I is a one-dimensional intensity or a three-dimensional color image and F is a feature image of $W \times H \times d$ dimension extracted from I,

$$F(x, y) = \Phi(I, x, y), \tag{3.5}$$

where the function Φ can be any mapping such as intensity, color, gradient, filter response, and so on. If we denote the d-dimensional feature points in R, which is a rectangular region included in F (R⊂F), as $\{z_i\}_{i=1..S}$, the region R is expressed by the dxd covariance of feature points

$$C_R = \frac{1}{S-1} \sum_{i=1}^{S} (z_i - \mu)(z_i - \mu)^T, \tag{3.6}$$

where μ is the mean of the points. The diagonal entries of the covariance matrix represent the variance of each feature, and non-diagonal entries are the respective correlations. Oncel et al. [73] explained the advantages of using covariance matrices as region descriptor of an image in their research. Among them, the greatest motivation for using covariance matrices in this study is that the representation of a covariance matrix can suggest a natural way to fuse multiple features that might be correlated. Thus, they also described that a single covariance matrix obtained in one region is usually sufficient to match the region of different views and poses. They also introduced the expression method called integral images for quick calculation of the region covariances. The integral images are intermediate image representation used to quickly calculate the region sums. For an intensity image I, its integral image is calculated as

$$\text{Integral Image}\,(x', y') = \sum_{x \le x', y \le y'} I(x, y). \tag{3.7}$$

In our study, we used integral images as a tool for dimension reduction of the calculated covariance matrix. We derived a feature representation named COV-II, an integral image for the covariance of CS [74]. We normalized COV-II and eliminated the outliers with extreme values. We extracted 13 attributes in total from the CS and the COV-II, including entropy, centroid, central moment and basic statistics. A pseudo-code for second feature extraction method so far is given in Algorithm 3.2.

Algorithm 3.2 A Feature Extraction Method 1 Pseudo-Code
input A snoring sound S, frame rate Fr, frame duration D, average ROI coordinate set O_C, window scaler V
1: Calculate frame size Fsize: Fsize = Fr×D
2: Extract frame data, FD by Fsize: FD ← S[$Fsize$]
3: Cyclic resolution dalpha = $1/(\frac{\text{Fsize}}{\text{Fr}})$, Spectral resolution df = $1/(\frac{\text{Fsize/V}}{\text{Fr}})$
4: Calculate the cyclic spectrum (CS): magnitude $(S_x) = \text{CS}\ (FD,\ df,\ dalpha)$
5: Crop the magnitude for extraction the region of interest $S_x crop = S_x[O_C\{x_1, x_2, y_1, y_2\}]$
6: Calculate covariance matrix of $S_x crop$: $S_{x\_crop}\text{cov} \leftarrow \text{covariance}(S_x crop)$
7: Transform the $S_{x\_crop}\text{cov}$ to integral image (COV-II): $S_{x\_crop}COV - II = Integral\ Image(S_{x\_crop}\text{cov})$
8: Calculate statistics (Entropy, centroid, central moment, max, min, median) from $S_{x\_crop}\text{COV} - \text{II}$
output $S_{x\_crop}\text{COV} - \text{II}$

In this study, the whole process was implemented and tested using Matlab® R2015a (MathWorks, USA) based on Windows PC (Intel Xeon 3.3GHz, 16GB RAM, Windows 10 Pro).

Key scheme	**Setting value**	**Parameter setting**	
Attribute selection evaluator	*SVMAttributeEval*	**Complexity**	*1.0*
		Epsilon	*1.0E-25*
		Filter type	*Normalization*
		Tolerance	*1.0E-10*
Attribute selection search	*Ranker (ranks attributes by their individual evaluations)*	**Generate ranking**	*True*
		Number of attribute to retain	*60*

Table 3.2 Parameter setting for feature selection in WEKA

3.2.3 Feature selection

Using the intermediate feature set which was obtained through the above processes, we set a goal of development of three types of a snoring classification model. In this study, WEKA toolbox [75], widely used the software in a data mining area, was used for the classification tasks. First of all, redundant and irrelevant features in our feature dataset were eliminated using the feature selection function to generate a more informative feature set and improve the classification accuracy. We performed the feature selection in the Algorithm 3.1. The output of feature extraction Algorithm 3.1 has 149 PCA coefficients as an intermediate feature set. To improve the classification performance in this task, we used an SVM attribute evaluator which was offered by the WEKA and could get a list of feature rankings. The top 60 attributes were finally selected from the ranking, and it became the final selected feature set. The parameter setting in WEKA related to the feature selection used in this experiment is summarized in the following Table 3.2.

In the case of Algorithm 3.2, not only was its number of output small but also there was no reasonable performance improvement when it applied to the

Index	Attribute	Remark
1	Max1	Maximum (standard deviation of Cov-II)
2	Mean1	Mean (standard deviation of Cov-II)
3	Median1	Median (standard deviation of Cov-II)
4	Max2	Maximum (mean of Cov-II)
5	Mean2	Mean (mean of Cov-II)
6	Median2	Median (mean of Cov-II)
7	Entropy	Entropy (Cov-II)
8	Central moment	Central moment (Cov-II)
9	Sx_sum	Sum (max. for magnitudes of CS)
10	Sx_entropy	Entropy (max. for magnitudes of CS)
11	Sx_max	Maximum (max. for magnitudes of CS)
12	Sx_centroid1	Centroid (max. for magnitudes of CS)
13	Sx_centroid2	Centroid (max. for magnitudes of CS)
Cov-II means the $S_{x\_crop}COV-II$		

Table 3.3 Attributes used in the final feature set

output, so the additional feature selection process did not perform. The attributes used in the final feature set are shown in the following Table 3.3.

3.2.4 Classification of three types of snoring

Using the final selected feature set, we tried to apply it to various machine learning algorithms and aimed to choose the most outstanding classifier. In the first experiment with 60 selected features, we chose three classifiers (SVM, Logistics, Random forest) which were embedded in WEKA and performed 10-repeated 10-fold cross-validation for each classifier. Finally, we analyzed the classifiers' performance through the corrected paired samples t-test. The paired

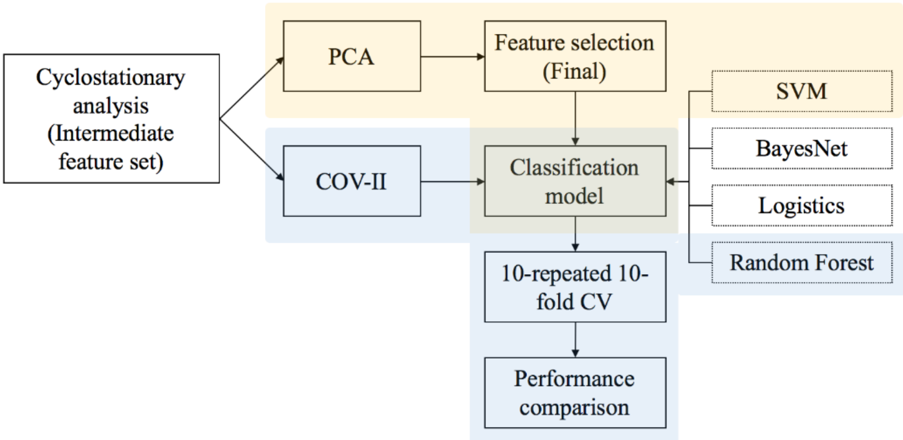

Figure 3.4 Process for the classification and its performance comparison

samples t-test is a statistical procedure to determine whether the difference between the mean of the two observations is zero. In this experiment, we used this procedure for confirming difference of average classification performance between each classifier.

In the second experiment with COV-II features, we selected a random forest classifier with 30 trees to confirm the classification performance through a classifier accuracy comparison test as in the previous experiment. This process of experiments is summarized in Figure 3.4.

3.3 Evaluation

In this study, we generated a classification model identifying three types of snoring based on statistics of cyclostationary features from the recorded breathing sound during PSG and validated its performance. We calculated the classification performance of different classifiers, SVM, Bayes network, multinomial logistic regression, random forest using 10-repeated 10-fold cross-validation.

<table>
<tr><th>Classifier</th><th>Key scheme</th><th colspan="2">Option</th><th>Remark</th></tr>
<tr><td rowspan="4">SVM[1]</td><td>Build LogisticModels</td><td colspan="2">False</td><td>Whether to fit logistic models to outputs</td></tr>
<tr><td>Complexity</td><td colspan="2">1.968</td><td></td></tr>
<tr><td>Epsilon</td><td colspan="2">1.0E-12</td><td>The epsilon for round-off error</td></tr>
<tr><td>Filtertype</td><td colspan="2">Normalize training data</td><td>Data transformation</td></tr>
<tr><td rowspan="4">BayesNet[1]</td><td rowspan="2">Estimator</td><td colspan="2">SimpleEstimator</td><td>Estimating the conditional probability tables of a Bayes network</td></tr>
<tr><td>Alpha</td><td>0.5</td><td>Initial count for estimator</td></tr>
<tr><td rowspan="2">Search algorithm</td><td colspan="2">K2</td><td>Bayes learning algorithm:
hill climbing algorithm</td></tr>
<tr><td>Score type</td><td>Bayes</td><td>The measure used to judge the quality of a network structure</td></tr>
<tr><td rowspan="2">Logistic[1]</td><td>ridge</td><td colspan="2">1.0E-8</td><td>Ridge value in the log-likelihood</td></tr>
<tr><td>maxlts</td><td colspan="2">-1 (unlimited)</td><td>Maximum number of iterations to perform</td></tr>
<tr><td rowspan="3">Random Forest[2]</td><td>maxDepth</td><td colspan="2">0 (unlimited)</td><td>Maximum depth of the trees</td></tr>
<tr><td>numFeatures</td><td colspan="2">0 (unlimited)</td><td>Number of attributes to be used in random section</td></tr>
<tr><td>numTrees</td><td colspan="2">60</td><td>Number of tree to be generated</td></tr>
<tr><td>Reference</td><td colspan="4">1: Setting value of Algorithm 3.1
2: Setting value of Algorithm 3.2</td></tr>
</table>

Table 3.4 key scheme options of each classifier used in this experiment

Cross-validation is a procedure of evaluating a prediction model by dividing the original sample data into a training set for model learning and a test set for evaluation of the generated model. Generally, in k-fold cross-validation (in our experiment, *k=10*), the original sample data is randomly divided into k equal-

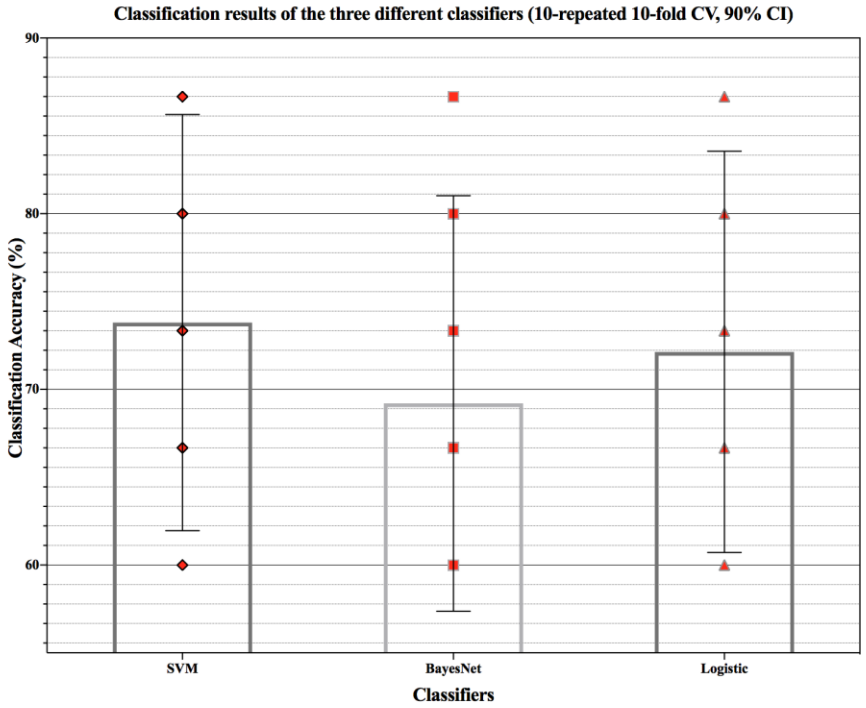

Figure 3.5 Snoring type classification performance of the three classifiers

sized subsamples. At this time, k-1 subsamples are used for learning data and the other one is test data for validating the model. Then, the cross-validation is repeated k times and in each process, k subsamples which are individual test sets, are used as validation data. Finally, a single estimation is calculated by averaging the results derived from each test. In the case of repeated cross-validation, the procedure is repeated n times (in our experiment, *n=10*) and producing n random partitions of the original sample data. From this, a single averaged estimation is yielded from n results. The key scheme options of each classifier used in this experiment are summarized in the following Table 3.4.

Detailed mean result (10-repeated 10-fold CV)	**SVM**	**Bayes Network**	**Logistic**
Percent correct (%)	73.80	69.20	72.13
Percent incorrect (%)	26.20	30.80	27.87
Kappa statistic	0.61	0.54	0.58
Mean absolute error	0.30	0.21	0.19
Root mean square error	0.38	0.43	0.41
Relative absolute error	67.17	47.04	41.89
Root relative squared error	81.04	90.38	87.33
True positive rate	0.82	0.66	0.80
False positive rate	0.19	0.21	0.18
True negative rate	0.81	0.79	0.82
False negative rate	0.18	0.34	0.20
F-measure	0.75	0.63	0.74
Area under ROC	0.84	0.81	0.89
Area under PRC	0.70	0.74	0.83

Table 3.5 Result summary of three classifiers

3.4 Results

3.4.1 Cyclostationary feature extraction using PCA (Algorithm 3.1)

The accuracies of these classifiers are 67.64 to 80.00 %, 63.04 to 75.35 %, 66.18 to 78.08% respectively with 90% confidence interval (Figure 3.5). Also, we conducted corrected paired t-test to compare the classification performance of classifiers statistically. As a result, there is no significant difference in percent correction among foregoing three classifiers. However, in comparison of the Area Under ROC (AUC), Bayes Network is significantly worse than other two classifiers. Therefore, SVM and Multinomial logistic regression are proper classifiers for classifying the symptom-related snores using the proposed cyclostationary features. Table 3.5 shows detailed mean results of 10-repeated 10-fold cross-validation of each classifier.

AUC range	Discriminatory abilities
AUC $\geq$ 0.9	Excellent
0.8 $\leq$ AUC $<$ 0.9	Good
0.7 $\leq$ AUC $<$ 0.8	Fair
AUC $<$ 0.7	Poor

Table 3.6 Criteria for clinical diagnostic testing based on the AUC value

In this thesis, the classifier performance is verified using the classification accuracy (Percent correct), sensitivity (True positive rate; TPR), specificity (True negative rate) and AUC. The accuracy represents the rate at which the analysis results of each sample are classified into the correct class. This measures the extent of veracity of the classification test on a condition [76]. The sensitivity value indicates the probability that the classification test identifies the patient suffering from the disease. Tests with high sensitivity tend to detect all possible positive conditions with fail and are useful for determining disease. The specificity is the proportion correctly identified as true negatives through classification results. This indicates how well the normal condition can be determined. In the ROC space, a TPR and a false positive rate (FPR) appear as a single point. TPR is equal to the sensitivity and FPR is equivalent to the specificity. The location of a single point indicates the tradeoff between the sensitivity and specificity, and shows whether classification performance is good or not. If the positions of points on the ROC are close to the diagonal, the accuracy of test becomes lower. The AUC is calculated by the integral of the ROC (i.e. area) and used as a measure of accuracy in clinical diagnostic trials [76]. Higher AUC values are considered to indicate better classification abilities. Table 3.6 shows the criteria for clinical diagnostic testing based on the AUC value [77].

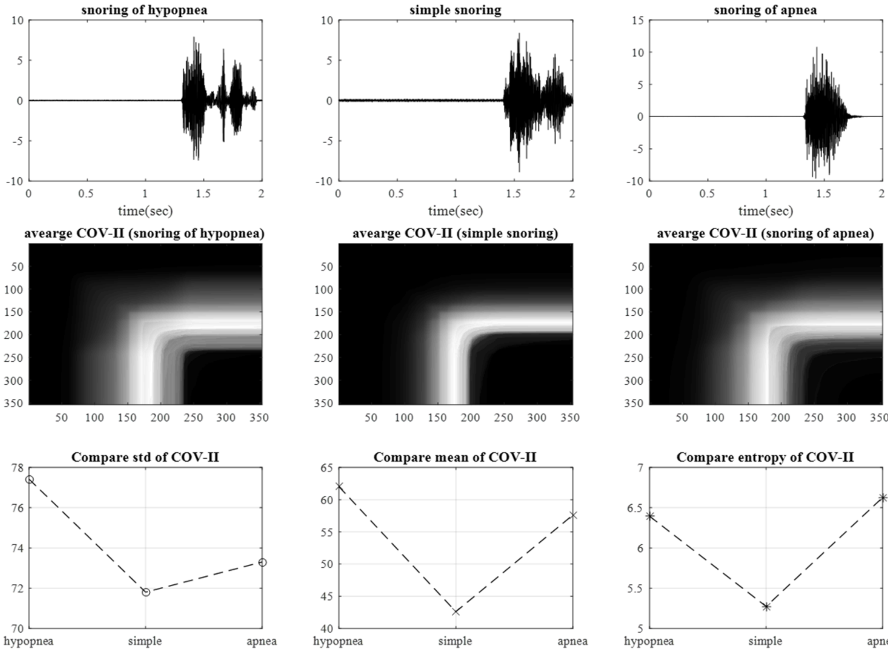

Figure 3.6 Integral image of the cyclostationary attribute's covariance matrix (from top to bottom: snoring events, average COV-IIs, and statistics of average COV-IIs)

3.4.2 Cyclostationary feature extraction using integral image of a covariance matrix (Algorithm 3.2)

Figure 3.6 shows that the average COV-II of 150 snoring sounds according to each event group (middle row). The nonzero region of each snoring event reveals different distribution and intensity. For instance, snoring of hypopnea (left) has highest standard deviation and meanwhile snoring of obstructive apnea (right) event shows the highest entropy. However, simple snoring (middle) has small values than other events across the board. This result could be interpreted that the proposed COV-II is useful to identify the snoring types using sound only.

We conducted the 3-class snoring classification task evaluation using four classifiers (Figure 3.7). The basic evaluation framework was the same as

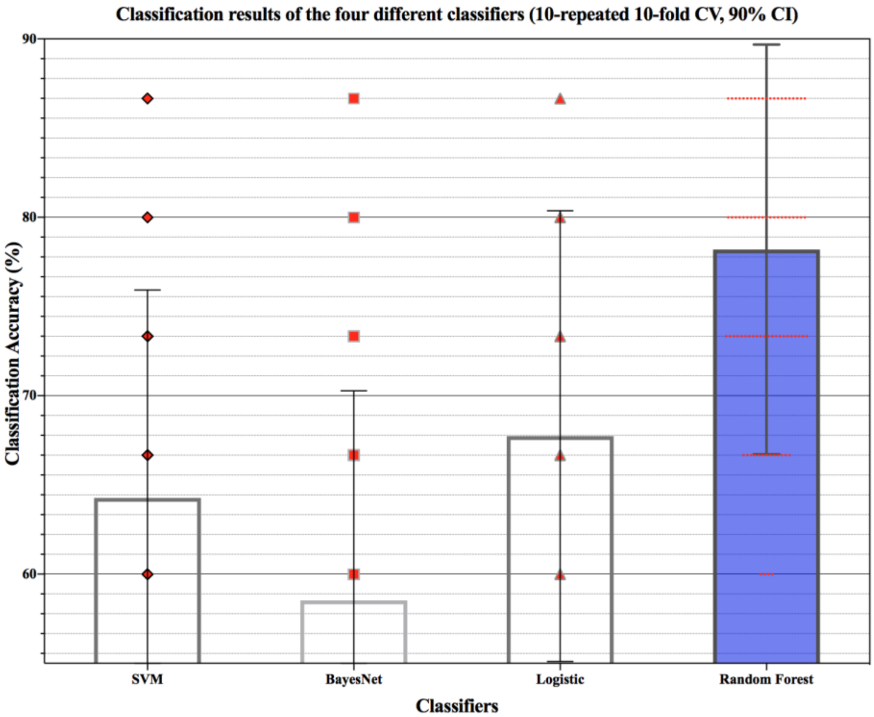

Figure 3.7 Classification results of the four different classifiers

preceding Algorithm 3.1. In the experiment using features derived from Algorithm 3.2, the random forest classifier showed the best classification performance. As shown in the above table, we conducted a three-class classification task using a random forest classifier with 30 trees, and other parameters were the default value of WEKA.

10-repeated 10-fold cross-validation was performed and the classification accuracy was 78.07±6.63% with a 95% confidence interval. Figure 3.8 shows the ROCs for each class. It is confirmed that AUC value is 0.8 or more for each SRBD-related snoring classification task. The analysis, comparison, and significance of two algorithms used in this experiment are described in the conclusion. In particular, both algorithms change the snoring sound to CS using cyclostationary analysis. However, the CS treats the two algorithms differently

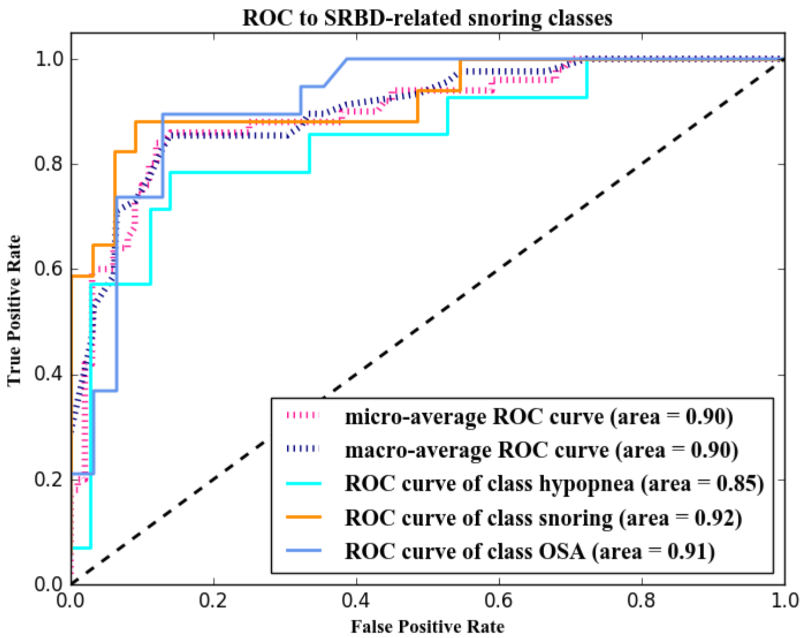

Figure 3.8 ROC to SRBD-related snoring classes

to generate individual feature representations. Therefore, we will focus on this.

3.5 Discussion

From the results of Algorithm 3.1, cyclostationary analysis can be used to estimate SRBD severity as features representing traits of three types of snoring which are observed during sleep. In the representative snoring signals shown in Figure 3.3, the nonzero values of the CS are distributed at various locations of the SCM converted from the snoring sound. By definition, a certain signal is cyclostationary when the nonzero magnitudes are exhibited for the nonzero values of alpha. These cyclostationary properties are presented with various attributes as a different magnitude and occurrence location. These attributes may be related to participant's upper airway characteristics which are involved in generating various SRBD related snoring sounds. In the case of simple and hypopnea related snoring in Figure 3.3, obvious lines depicting high spectral correlation appear in the alpha frequency domain along the frequency domain. Furthermore, simple snoring showed the higher magnitude and relatively high magnitude regions are more widely located than the hypopnea-related snoring. On the contrary, in the case of the OSA-related snoring (special loud snoring after the long cessation of breathing during sleep), nonzero spectral correlations are widely distributed, and a high magnitude area is also presented more widely than the other two types of snoring. Due to many factors influencing snoring sounds of patients, for example, recording environment and the position of patients, distributions, and magnitudes of components do not show definitely same patterns. However, by using simple statistics and PCA, we could extract approximated compact features based on the cyclostationarity related to three types of snoring.

In the case of Algorithm 3.2, we obtained the covariance matrix from the CS and applied an integral image method to it. From the feature representation applying the integral image, we extracted various statistical values, and also added other simple statistical values based on original CS to generate the final feature set. In particular, this final feature set showed better classification performance without the feature selection process, unlike the Algorithm 3.1. This means that the attributes of Algorithm 3.2's final feature set is less redundant and irrelevant than Algorithm 3.1. Algorithm 3.2 derived the covariance matrix from the original CS. In this experiment, the CS is treated as a 2-D image of the contour view type as shown in the previous Figure 3.3. Since covariance of an image is known to play a role as an object descriptor in the field of image processing, it is considered to be more efficient feature set than the product of Algorithm 3.1. We obtained the covariance matrix from the magnitude values of the CS derived using the cyclostationary analysis. The original CS has a dimension of $(32{,}001 \times 401)$ in our experiment. Its transposed matrix was able to generate a (401×401) dimensional covariance matrix. This procedure can also be considered as a dimension reduction process. In this case, the variables that obtain the covariance can be defined as X_n $(n = 1,2, \ldots 401)$ composed of 32,001 CS magnitude values. The diagonal entries of the covariance matrix represent the variance for each X_n and the remaining entries represent the covariance values for each pair of different X_n. Positive covariance between two variables indicates a positive linear relationship, and negative covariance indicates vice versa. A covariance value of zero indicates that there is no linear relationship between the two variables and that they are independent of each other. In this study, X_n is a variable including α frequency values (the CS magnitude values) belonging to one spectral frequency value. Therefore, the covariance matrix consists of variance values of X_n, and the covariance values between the CS magnitude values that correspond to each spectral frequency value. For example, if the covariance values of X_1 and X_2 are positive, the CS magnitude values included in these values exhibit a positive linear relationship, that is, when X_1 increases, X_2 also shows an increasing characteristic. We also applied integral images to the covariance matrix. As explained earlier, in the integral image, all pixels are the sum of the upper and left pixels, and consequentially, it can calculate the summations for the subregions of the image. As can be seen in the result, the statistical values of the integral images were different according to the types of snoring. Therefore, when Algorithm 3.1 and 3.2 are compared based on the necessity of the feature selection process, it can be advantageous to grasp the correlation and summarize related values rather than to use the primary result of the cyclostationary analysis to derive a more useful and reliable feature set.

In two experiments using different features, we compared the classification performance by different classifiers. Experiments using the algorithm 3.2 with better classification accuracy showed the best performance when using a random forest. A random forest is an ensemble method that combines bagging with tree models that are sensitive to training data fluctuations. Bagging is a simple, yet effective ensemble method for generating various models for different random samples of a raw data set. When bagging is applied to a tree model, a process called subspace sampling, which constructs individual trees from different random subsets of features, is performed. The decision tree is a grouping model in which leaves partition the instance space, which is a collection of features, so a corresponding instance space partition of the random forest is the intersection of the partitions of the individual trees in the ensemble. Therefore, the random forest partition is more delicate than most tree partitions, and the random forest is an alternative learning algorithm for the tree models because the ensemble has an effective decision boundary that cannot be learned by the single base classifier [78]. The features output through Algorithm 3.2 consists of statistical values based on CS and its integral image of covariance matrix. As shown in Figure 3.6, the tree model can be a good solution when performing classification tasks because each statistic value has different values according to the related symptoms. Therefore, it is considered that the best classification performance is obtained from a random forest having a finer partition and a more effective decision boundary than the general tree models.

Another finding of this study is about the possibility of hypopnea-related snoring identification. The typical apnea-related snoring contains relatively obvious information in a time domain, which represented by a loud snoring sound occurrence after a long-term breathing stop. However, hypopnea snoring was not easy to compare with simple snoring sound just by listening to the sound and also it did not have certain features in the time domain as OSA-related snoring. However, hypopnea snore was not easy to compare with simple snore just by listening to the sound, and also it did not have certain characteristics related to sound generation in the time domain like the OSA-related snoring. Besides, hypopneas are typically identified by the bio-signal indicators, such as the oxygen saturation, nasal pressure transducer or an electroencephalographic arousal signal. However, the fact that hypopnea-related snoring can be distinguished using the breathing sounds brings us a possibility that cyclostationary analysis may be adopted as one of the main features for designing a sleep disorder severity classifier related to SRBD.

3.6 Summary

In this experiment, we extracted three types of snoring sounds based on the annotations of the physician's SRBD-related snoring (snoring, hypopnea, OSA) from recorded sleep breathing sounds during the PSG test. Although there have been a variety of related studies, very few studies have attempted to distinguish hypopnea related snoring using only breathing sounds. Unlike OSA, hypopnea is not clear about its recognition criteria, and it mainly uses EEG or oxygen saturation information, which is not a characteristic of sound or time in the process of recognition. Therefore, in this experiment, it is crucial how accurate the symptomatic snoring sounds can be extracted by the gold standard. Unfortunately, we have not been able to find any experimental frameworks or databases that have been created and used as Gold Standard for snoring sound analysis. Therefore, we used the annotation information of the physician set as the time interval as the gold standard and extracted the sound data of the matching interval according to the symptom.

For the three snoring sounds, this experiment showed a classification performance of over 70%, and the AUC at that time was more than 0.8, which shows this method is useful in clinical practice [77]. However, as mentioned earlier, there are obvious limitations because data extraction for individual events is difficult and the number of data used in experiments is small. In the future, it will be necessary to increase the number of data and develop a robust classifier that can cope with various conditions through the development of a clear snoring sound event extraction framework based on the physician's annotation.

However, this experiment showed that cyclostationary analysis could identify the snoring sounds by symptoms. This means that this analysis method can be used for snoring event classifiers and can extract various features when applied to the sleep breathing sound analysis. In the next section, we conducted an OSA severity classification study by applying this analysis method to the long-term sleep breathing sound of individual patients.

Chapter 4. Patient's OSA Severity Classification

In this thesis, we perform two kinds of sleep breathing sounds analysis tasks. The first is to distinguish SRBD-related snoring events from snoring sound units extracted from full-night sleep breathing sound described in previous chapter, and the second is to obtain a specific feature representation from long-term sleep breathing sound and use it to classify OSA severity of individual patients. This chapter describes the second task.

4.1 Introduction

As mentioned in Chapter 1, obstructive sleep apnea (OSA) is the most common sleep-related breathing disorder. Chapter 4 is a study that directly distinguishes OSA severity by using sleep breathing sounds. After explaining the more detail OSA-related content, we will explain the actual experiment process and the result.

The OSA syndrome is characterized by repetitive episodes of upper airway obstruction and commonly connected with a reduction in blood oxygen saturation. OSA is associated with a characteristic snoring pattern and consists of loud snores or short gasps that alternate with events of silence that typically last for more than 10 s. OSA also can induce various dangerous situations or personal complaints in person's daily life. For example, the patient's severe daytime sleepiness due to OSA can be a direct cause of a large number of car accidents [79]. Besides, gastroesophageal reflux can occur as a result of the effort made to reestablish breathing [80], and the loss of both libido and erectile ability could occur in patients with OSA [81]. Cardiac arrhythmias also commonly occur during sleep in OSA patients [82]. In this case, bradycardia alternates with tachycardia during the apneic phase and termination phase of the obstruction, respectively. Even more severe, tachyarrhythmias most commonly occur when a patient tries to reestablish breathing following the apneic phase and may increase the risk of sudden death during sleep.

Many population-based studies have reported a high prevalence of OSA in adults [83]. In the case of the United States, OSA has increased over the past two decades and its prevalence rate in adults between 30 and 70 years old has reached 26 % [84]. Despite the seriousness and increased cases of OSA, related research has reported that 93 % of women and 82 % of men remain underdiagnosed [85]. The main reason for the high number of underdiagnosed individuals is that it is hard to recognize the intensity of their pathological breathing during sleep. Even if they are aware of the symptoms, an expensive and uncomfortable examination make their visit to the hospital difficult.

Polysomnography (PSG) is currently the gold standard for the diagnosis of OSA. To make an OSA severity diagnosis, PSG provides an Apnea-Hypopnea Index (AHI) that contains the number of apnea and hypopnea occurrences per hour of sleep. According to the American Association of Sleep Medicine (AASM), when a subject has more than five obstructive apneas over 10 s per hour of sleep, the individual could be suspected of having OSA syndrome [86]. However, the test should be conducted overnight, and its cost is expensive. Moreover, the measurement is inconvenient because various physiological sensors must be attached to the body [87]. Because of these limitations of PSG, it is not suitable for mass examination and occasionally, obtaining reliable results is difficult due to different sleeping behaviors in a hospital, called first night effect (FNE). Recently, portable PSG has been developed and used in personal home care related to sleep disorders. However, this technology still requires multiple uncomfortable sensors and measurements for various physiological parameters, such as blood saturation and nasal airflow. Therefore, a preliminary screening test is necessary for suspected subjects who are concerned about the financial burden and measurement inconvenience. The test should be as simple as possible and should be capable of repeatedly measuring patient's data for mass examinations.

Breathing sounds can be measured more easily than other known physiological signals during sleep. Conventional sensors can be used to take measurements in a body-contact manner, but the breathing signal can be recorded using non-contact sound recording devices. If the recording device is not professional and the operation is not complicated, the breathing sounds may be measured without any help from specialists or technicians. Decisively, many studies show that sleep breathing sounds are related to sleep disorders [17, 45, 57, 58, 62-64, 67]. These studies could be representative examples of the medical advantages of being able to examine some symptoms related to sleep disorders without additional bio-signal sensors when high-quality sleep breathing sounds can be obtained from patients. Therefore, sleep breathing sounds can be regarded as acoustic physiological signals that anyone can measure. However, most recent studies have focused on snoring segment detection, snore/non-snore classification, or OSA/non-OSA patient group classification. The sensitivity result of OSA classification, which has shown that a percentage of people with OSA are correctly identified as having the symptom, ranges from 60 % to 80 % in related studies [45, 62, 63, 68]. For an efficient OSA screening test, OSA severity should be able to report results based on a clinical standard. According to the AASM, AHI values are categorized into four severity labels: normal, mild, moderate, and severe sleep apnea. Moreover, many studies have used body-contact microphones, for example, microphones attached to a surrounding area of the neck [17, 62, 64] or face [63]. These contact microphones easily cause inconvenience for patients and make it difficult to make simple measurements. Additionally, numerous studies have acquired breathing sounds using expensive professional microphones that typically hang from the ceiling at a short distance from the patient. More detailed information on the algorithms of previous studies is presented in a discussion section comparing the results of other studies with those of the proposed study. The aim of this study is to develop a new approach to multiple OSA severity classification using breathing sounds during sleep. Two novel methods, the total transition probability of approximated sound energy in a time series and the statistical properties which are derived from dimension-reduced cyclic spectral density, are proposed for our object. To the best of our knowledge, so far, no approach has utilized a combined feature set, which was made with prior methods, for multiple OSA severity classification. In contrast to related studies [17, 62-64], breathing sounds are recorded using an ordinary microphone that is placed at a long distance from the patient and not intended to record special sounds, such as the patient's breathing. Moreover, we focus on breathing sounds during non-rapid eye movement (NREM) sleep: sleep stages 2 and 3. We know that sleep apnea-related snoring is most likely to occur during REM sleep. However, because we use an ordinary subject's breathing sounds, we concentrate on the two sleep stages in which conventional snoring is most likely to occur. Furthermore, body movement or other complex behaviors rarely occur during these stages, hence we can minimize the noise that is unrelated to breathing sounds. Additionally, we attempt to extract succinct characteristics from relatively long audio recordings without any particular event detection method or random event selection in contrast to previous studies [17, 58, 62, 64, 67].

4.2 Existing Approaches

There have been many studies on OSA detection using breathing sounds during sleep. In this section, we summarize some studies that are considered to be highly relevant to this study. Looking at the research methods of various studies, it can be seen that most of the experiments used professional recording environment and special microphones. It is the biggest difference from our previous studies that we did not consider setting up a special recording environment in our experiment.

Nakano et al. [62] recorded the tracheal sound using a body-contact microphone and calculated a transient fall (TS-dip) of the power spectra's moving average in the time series. With this feature, they obtained the tracheal sound-respiratory disturbance index (i.e., the number of TS-dips per hour) and compared it with existing AHI values from PSG. The result of OSA subject detection using their feature (AHI threshold 5) was 93 % sensitivity and 67 % specificity. Abeyratne et al. [45] detected segments of snore related sounds (SRS) detected automatically and categorized SRS into pure breathing, silence, and voiced/unvoiced snoring segments. From these segments, they extracted the intra-snore pitch periods feature, which was characterized by discontinuities called intra-snore-pitch-jumps. Using this feature, they obtained an OSA detection result with 100 % sensitivity and 50 % specificity, where the AHI threshold was 5. Azarbarzin et al. [64] recorded sleep breathing sounds with a special microphone that was placed over the suprasternal notch of the trachea, and extracted three types of the segment: non-apneic, hypopneic, and postapneic. From these segments, they calculated the total variation norms of the zero-crossing rate and peak frequency and used them as features. They obtained 77.2 % accuracy from the four-OSA severity classification test and an additional OSA detection result of 92.9 % sensitivity and 100 % specificity with AHI threshold 5. Behar et al. [63] detected OSA subjects using breathing sounds and additional information from sensors such as actigraphy, body position assessment, and photoplethysmography (PPG). For breathing sound recording, they used a special microphone attached to the subject's face and extracted the multiscale entropy values from the audio. The OSA subject detection result based on audio was 69.5 % sensitivity and 83.7 % specificity for training using SVM.

4.3 System Architecture

To develop an OSA severity classification method using breathing sound, individual respiration sounds were divided into four OSA severity groups. All breathing sounds were acquired from video clips for PSG room monitoring that were included in the clinical sleep diagnostic tool.

In the preprocessing, the spectral subtraction technique which is popular for the enhancement of noisy speech signal was applied [69]. Then, the total transition probability of the approximated sound energy and the statistical properties of the cyclic spectral density features were extracted from the preprocessed breathing sounds. These features and machine learning techniques were used to train the OSA severity classification model and verify its accuracy.

In this section, we first describe participant statistics, physical recording environment, and sound acquisition method. Secondly, we explain the details

Parameter	Conditions	SURP-102 (PSG room)	SPH0644LM4H-1 (Samsung Galaxy S8)
Directivity		Omnidirectional	
Supply voltage		9 ~ 12 V	1.6 ~ 3.6 V
Frequency range		20 ~ 2k Hz	100 ~ 10k Hz
Sensitivity	94 dB SPL @ 1kHz	-40 dB	-37 dB (typical)
Signal to Noise Ratio	94 dB SPL @ 1kHz, A-weighted	60 dB	65.5 dB

Table 4.1 Comparison of microphone used in experiment with latest smartphone

of the feature extraction methods. Finally, we describe the training and validation method for the classification model.

4.3.1 Breathing Sound Database

A total of 83 adult subjects (27 females and 56 males with a mean age of 48.7(±17.5) years, mean body mass index of 25.6(±4.1), and mean AHI of 23.6(±25.3)) were enrolled from the sleep laboratory of the Seoul National University Bundang Hospital (SNUBH), South Korea. The study was approved by the institutional review boards at SNUBH, and informed consent was obtained from all patients or their guardians on their behalf.

The basic environment of the laboratory is the same as previous Chapter 3. Again briefly, the PSG room contained a video camera and auxiliary microphone (SURP-102, YIANDA electronics Co., Ltd, ShenZhen, China; 20-2kHz frequency range, -40dB sensitivity) for monitoring the test for the entire night. The video clips were synchronized with various physiological signals of

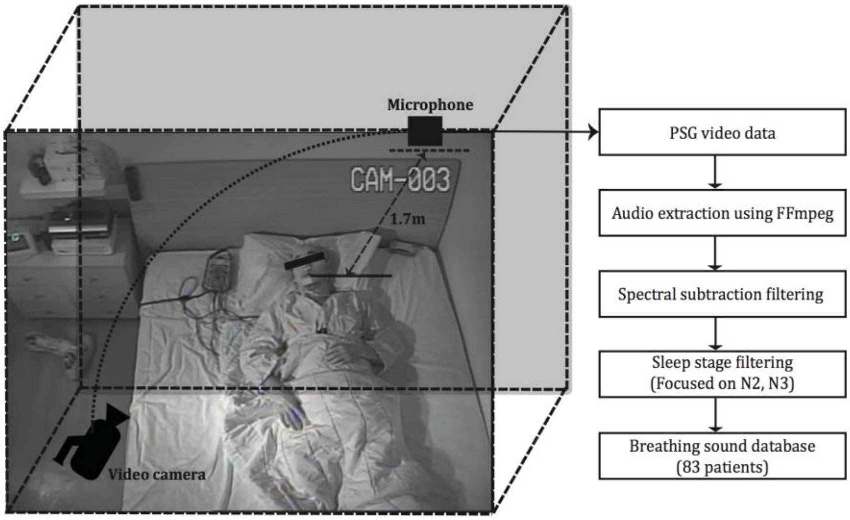

Figure 4.1 Sound acquisition and preprocessing in the PSG room

the PSG and stored through the sleep laboratory's sleep diagnostic software (REM-Logic, Natus Medical Inc. CA, USA). The auxiliary microphone was set up by default in the PSG room and was not installed for clinical purposes. It was located on the ceiling 1.7 m from the bed and has a frequency range of 20 ~ 2kHz according to its device specification. In general, the minimum frequency bandwidth required to carry human voice is known as 300 to 3,400Hz. This corresponds to the corded telephone voice frequency band (VFB). The microphone used in this experiment has a narrower frequency range than VFB and has a lower frequency range than VFB, but does not reach in the high-frequency range. This can be just like applying a 2kHz low-pass filter to a high-performance microphone with a frequency range of 50 to 20kHz. Since it was hard to obtain a full specification of the microphone used in our experiment, it was practically difficult to compare it with other kinds of microphones directly. However, it is evident that its performance is lower than the microphones used in the latest smartphones. The following table compares the microphone utilized in the most recent smartphones with that used in our experiment.

In conclusion, it is highly probable that the recorded sleep breathing sounds in this situation are not as good as the sounds obtained through well-controlled professional recording environment as in previous studies. Although the position of the microphone on the ceiling is not a very common situation from the user's point of view, we can be confident that this experiment is based on a relatively general recording environment rather than other research. Therefore, we aimed to develop a classification algorithm that can distinguish OSA severity based on relatively simple and general environment that can record the sleep breathing sounds through an ordinary microphone which is carelessly suspended from a ceiling.

The left-hand side of Figure 4.1 shows the actual setup of the PSG room. From monitoring-video clips of the PSG room, the all-night breathing sounds of each subject were extracted using a multimedia conversion tool (FFmpeg) [68] and saved as a wave format file with an 8 kHz sampling frequency. Then, according to each patient's AHI value from the PSG test result, wave files were categorized into four OSA severity groups: normal ($0 \leq$ AHI ≤ 4), mild ($5 \leq$ AHI ≤ 14), moderate ($15 \leq$ AHI ≤ 29), and severe (AHI ≥ 30). The normal group included 20 breathing sounds and all other OSA groups contained 21 sounds. The average time of the sleep breathing sounds was 7 hours 10 minutes 30 seconds.

4.3.2 Preprocessing

Since the recording equipment was not specifically chosen for sound analysis experiments and was intended to monitor the PSG room's test environment, there was a lot of background noise, such as white noise, hum or hiss when we checked the audio signals. We considered that these noises' spectrums did not substantially change the target signal; thus, we adapted a spectral subtraction method [69], which is a computationally cheap and efficient method for this situation. We assumed that subject's regular breathing sounds during sleep could be used to estimate OSA severity. When we identified a typical hypnograms of adults, NREM sleep corresponding to stage 2 and 3 accounted for at least 60 % of the total sleep [88]. During these sleep stages, the subject is typically stationary and the respiratory pattern is regular. Therefore, we assumed that regular breathing sounds could be obtained and various noise associated with the subject's body movement and arousal could also be minimized during these stages. Thus, we extracted sounds corresponding to the stage 2 and 3 NREM sleep in the original breathing sound database. Breathing sound data were simultaneously stored with various physiological data from the PSG test and synchronized. In addition, sleep stages were labeled by sleep specialists or physicians using the clinical sleep diagnostic software mentioned previously. Such label information can be exported to the outside of the system by applying various filtering in the program. Therefore, we could extract breathing sounds based on the time-stamped sleep stage information from the software.

The average time of all extracted breathing sounds related to stages 2, and 3 NREM was 4 hours 1 minute 55 seconds (± 1 hour 34 minutes 59 seconds), which was significantly reduced sound data compared with the original sound database. The right-hand side of Figure 4.1 shows the preprocessing procedure used in the present study. The breathing event detection method was not adapted to breathing sound analysis, in contrast to previous related studies [17, 58, 62, 64, 67]. This strategy can reduce possible detection errors and computational cost during the pre-detection process for target events. Whereas, our method simply divides the breathing sounds into window units of a predetermined window length and inputs them sequentially to the proposed feature extraction method. In this approach, most respiratory sounds associated with normal breathing, snoring and other disorders can be used for OSA severity estimation. Naturally, it is meaningful if the proposed method can extract representative features that can show respiratory characteristics in the individual signal window.

4.3.3 Feature Extraction Methods

We have considered two key features related to the time and spectral domain in this experiment. The features in the time domain have the advantage of being able to determine the temporal features of the OSA and other breathing events. Spectral features also provide valuable information such as the hidden attributes of each sound unit; thus, it could identify the target data that could not be recognized through the time domain features.

The first feature in this study is the total transition probability of approximated breathing sound energies in the time domain. The second feature is derived from the cyclostationarity-based information of breathing sounds which presents hidden spectral characteristics using the periodicity of the signal's autocorrelation. Then, this was simplified and transformed into a statistical representation. All features were calculated using signal processing and statistical functions of MATLAB 2015a (MathWorks, Inc., MA, USA) on a Windows PC (Intel Xeon 3.3GHz, 16GB RAM, Windows 10 Pro).

In the next sections, we will describe the previous two feature extraction methods: time domain and cyclostationarity based feature.

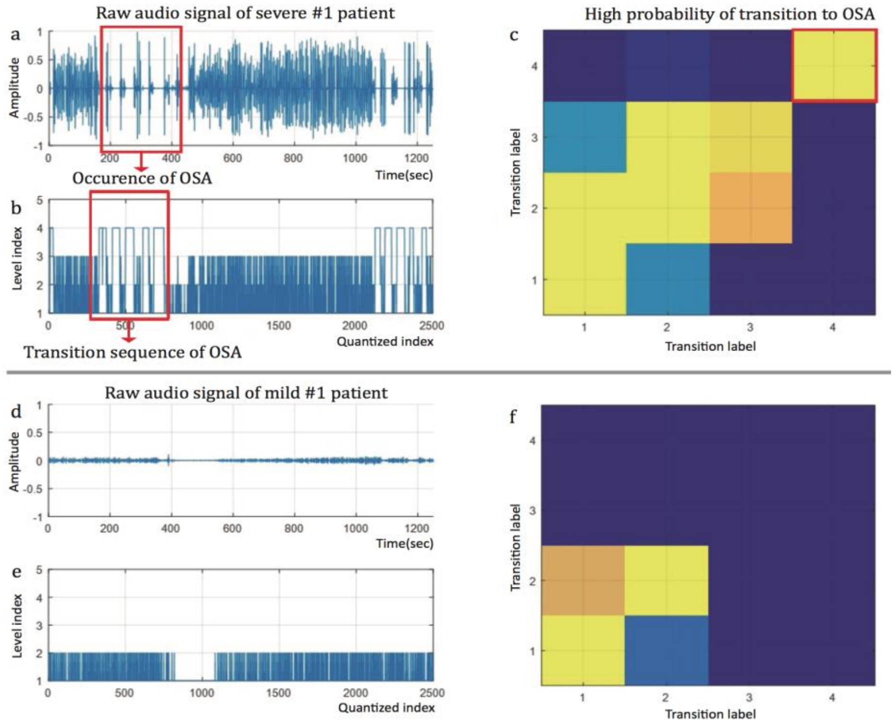

Figure 4.2 Representative example of the time domain analysis of sleep breathing sound.

a: A raw audio data example of the OSA severe group, b: its quantized signal's energy.

c: The 4 × 4 transition matrix including probability values, which is calculated by (b).

d–f: Example results of the OSA mild group

4.3.4 Time Domain Analysis

Sequential changes of the PSG data in the time domain are primary information that can be used clinically to diagnose sleep apnea. For example, certain episode in which respiratory arrest lasts longer than 20 seconds are an important indicator of sleep apnea [86].

We assumed that as the frequency of obstructive sleep apnea increased, the frequency and length of silent intervals between breathing sounds would be increased, so that the sequential amplitude changes of breathing sounds during sleep can represent the incidence of obstructive sleep apnea. Using this characteristic, we summarized the signal transition information of the subject's breathing sound as features for OSA severity classification. Using a 0.5 s Hanning window with 80 % overlap, each subject's breathing sound was segmented. Each segment was transformed into energy values and then approximated into three simple energy levels using thresholds (level 1: silence, level 2: lower energy level, and level 3: higher energy level).

Snoring has two dominant patterns of simple and complex waveforms. The complex-waveform snoring is associated with palatal snoring and may be indicative of actual airway obstruction through a collision of airway walls. The simple waveform snore does not actually obstruct the lumen, but it is generated by the oscillation around its neutral position and is associated with tongue-based snoring. The previous study shows that the palatal snoring has a higher ratio of peak sound amplitude to effective average sound amplitude than nonpalatal snoring [42].

Based on these facts, we assumed that level 2 can represent energy level of general breathing events, including simple snoring, and level 3 contains louder snoring's energy level associated with OSA events. Two dynamic thresholds were applied to divide the energy signal into two levels, which were sequentially updated from a predefined ratio of the most frequent energy peak range in each window segment. To eliminate some ripples, which were produced by an accumulation of the low energy values of the energy conversion process, we calculated another threshold, which was the proportion of actual signal energy within the area of a window. For instance, if the maximum energy value in a certain energy window was higher than the lower threshold and its energy proportion was more than 50 %, this window was simplified as level 2. On the other hand, if the energy ratio is less than 50 % under the same conditions, this window frame is approximated to level 1. The threshold was applied to the early energy conversion process and reduced the errors of the energy approximation process.

After changing the audio signal to three levels as described above, a second analysis process was performed. OSA suspected sections were searched using the length of level 1 (silence) and the occurrence of different levels (> 1) on both sides of the level 1 section. When the level 1 section lasted for more than 20 s and was located between levels 2 and 3, this silent section was changed to level 4, which was an extra level for the OSA suspected section; that is, level 4 particularly indicated that this section is an OSA candidate.

During the third stage of analysis, those mentioned above approximated and weighted signal was transformed into a transition matrix. Since the signal had four levels, a transition matrix form was 4×4, and the number of occurrence of 16 transition cases was accumulated in the matrix elements. By normalizing the matrix, the cumulative numbers of the elements were transformed to probability values, which represented the tendency of the subject's breathing sound energy transition. As a result, these 16 probability values were used as features of time domain analysis in this study. Figure 4.2 shows a representative example of the conversion results of this method. A pseudo-code for time domain feature extraction method so far is given in Algorithm 4.1.

Algorithm 4.1 Pseudo-code for time domain feature extraction method

input A snoring sound S, Sample rate in Hertz FS, Quantization level QL, Window W, Window length percentage WP, Overlap percentage OP

1: $FS \leftarrow 8000, QL \leftarrow 3, WP \leftarrow 0.5, OP \leftarrow 0.8, W \leftarrow Hanning$

2: Calculate window length WL and frame number FN:

$$WL = WP \times FS, FN = length(S) / WL$$

3: Signal framing with start point SP and end point EP:

$$SP = WL \times (k-1) - OP \times (k-1) + 1, EP = SP + WL - 1$$

where, k is window index (k = 1,2, ..., FN)

4: Calculate the energy in separate windows (E) and merge them into ES:

$$ES\ (Energy\ signal) \leftarrow E = energy\ (S(SP{:}\ EP), W, WL)$$

5: Search the most common max peak value in ES:

$$ES \leftarrow ES \leq mean(ES) + 6 \times std(ES)$$

$$peaks, location = findpeaks\ (ES)$$

$$make\ histogram\ using\ peaks{:}\ n, edges = histcounts(peaks)$$

$$find\ most\ frequent\ max\ peak\ value\ (FP) \leftarrow edges(max\ index)$$

6: Calculate threshold 1(TH1) and threshold 2 (TH2):

$$TH1 = FP \times 0.3, \qquad TH2 = FP \times 1.5$$

7: Calculate proportion of actual signal energy within the area of a window

$$Area\ ratio\ (AR) = {\int E}\Big/{WL \times FP}$$

8: Approximate energy at three levels:

$$Level1 = E < TH1$$

$$Level2 = E > TH1\ and\ E > AR \times 0.5$$

$$Level3 = E > TH2\ and\ E > AR \times 0.5$$

9: Find an OSA candidate section

$$Level4_{section(n)} \leftarrow if\ time\ of\ Level1_{section(n)} \geq 20\ s\ and\ \leq 40\text{, and}$$

$$if\ Level1_{section(n-1)} = Level2\ or\ Level3\ and$$

$$Level1_{section(n+1)} = Level2\ or\ Level3$$

10: Make a transition matrix using the Level section information

$$Transition\ matrix\ (TM) \leftarrow count(Level1, \ldots, Level4)$$

11: Converting a TM into a probability matrix (pTM):

$$pTM = normalization(TM)$$

output pTM

4.3.5 Cyclostationary analysis

As mentioned earlier, we did not consider the sound analysis method based on the breathing events detection. Therefore, the signal window, which is the basic unit of analysis, may contain unspecific waveforms or noise that are not related to snoring related sounds. Furthermore, the dominant signal in a window was considered as a nonstationary signal containing repetitive and complex waveforms such as breathing or snoring. Thus, characteristic properties representing not only the window's basic properties but also an overall summary were required for our goal. In the case of snoring, it is the result of an obstruction of air flows in the respiratory tract during sleep and causes repetitive vibrations of the tissues of the throat [49].

The snoring sound could be considered as including two main types of waveforms. The first is a complex waveform with a low-frequency sound that is generated as a result of the collision of opposing airway walls during passing periods of airway obstruction. The second is a simple waveform sound with a quasi-sinusoidal pattern that could be considered as a result of the airway walls' vibration around a neutral position without an obstruction of the respiratory tract lumen [89].

To obtain valuable insight of sleep breathing sounds in this experiment, we applied the cyclostationary analysis which is an unattempted method in this field. It is possible to deduce that a signal is cyclostationary when it is nonstationary, and its statistical characteristics vary periodically in the time domain [71]; that is, if the signals can decompose to the several sinusoidal wave components through a nonlinear transformation of order n, the signals are defined as an n-th order cyclostationary process.

In the present study, an autocorrelation function, a second order statistic, was used for the nonlinear transformation of the signals. Therefore, if the second-order statistic of the signal is periodic, it is second-order cyclostationary. We first calculated a bivariate autocorrelation function $C_{xx}(t,\tau)$ for each window of breathing sound $x(t)$ (time *t*, time lag *τ*). Then, we converted this function using a two-dimensional Fourier transform, and the result was the spectral density. The spectral density $S_{xx}(\alpha, f)$ must consist of two frequency variables: frequency f and cyclic frequency α. Only if the f and α frequencies were related to some hidden frequencies in the stochastic process, $S_{xx}(\alpha, f)$ is a continuous function of f and is a discrete function over α with non-zero values [90]. This non-zero spectral density, called the cyclic spectrum, and the spectral density $S_{xx}(\alpha, f)$ are commonly known as the cyclic spectrum (CS):

$$C_{xx}(t,\tau) = E\left[x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})\right], \tag{4.1}$$

where * denotes the complex conjugate,

$$S_{xx}(\alpha, f) = \lim_{T\to\infty}\frac{1}{T}\int_{\frac{-T}{2}}^{\frac{T}{2}} C_{xx}(t,\tau)e^{-j2\pi(f\tau+\alpha t)}dtd\tau. \tag{4.2}$$

To derive the overall cyclostationary features of a long-term sleep breathing sound, the sequentially updated mean of CS, $rmS_{xx}(\alpha, f)$ was calculated using the current CS and a previous mean value for every 60 seconds window, where $rmS_{xx}(\alpha, f)$ is described as follows:

$$rmS_{xx}(\alpha, f)(k) =$$

$$rmS_{xx}(\alpha, f)(k-1) + \frac{S_{xx}(\alpha, f)(k) - rmS_{xx}(\alpha, f)(k-1)}{k}, \tag{4.3}$$

where *k = 1,2,3, ..., N* and *N* is the total number of windows.

The number of steps in the f and α domains was 54 and 889, respectively. Therefore, $rmS_{xx}(\alpha, f)$ generated a 54×889 matrix, with the magnitudes of $rmS_{xx}(\alpha, f)$ as the elements. However, it was not appropriate to use the

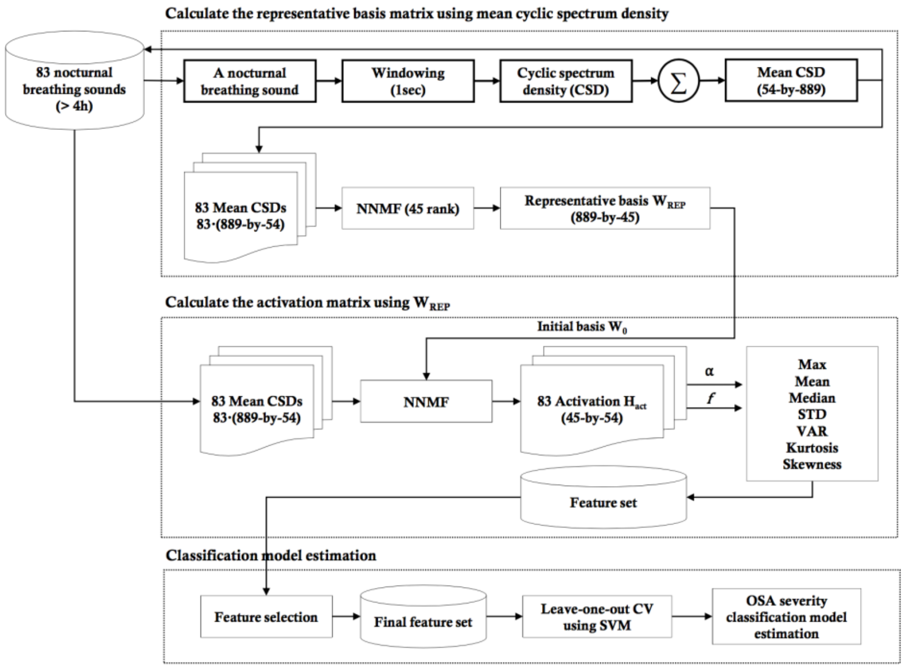

Figure 4.3 Feature extraction and classification based on nonstationary analysis. Statistical cyclostationary properties were extracted using the mean cyclic spectrum (CS) and the non-negative matrix factorization (NMF) for dimension reduction

complete matrix of $rmS_{xx}(\alpha, f)$ as a feature because of the widespread zero values. For dimensionality reduction, first, a threshold using Otsu's method [51, 91] was applied to the $rmS_{xx}(\alpha, f)$ matrix to eliminate unnecessary zero regions. Secondly, a non-negative matrix factorization (NMF) technique was applied to the previous matrix. NMF can analyze large quantities of data through the approximated decomposition of target data V into non-negative factors, which consist of a basis matrix W that includes inherent properties of data, and a coefficient or activation matrix H [92]. To obtain a representative basis matrix, a total of 83 previous $rmS_{xx}(\alpha, f)$ matrices were sequentially merged into an input matrix, and the basis matrix was calculated using NMF with rank 45. The rank was heuristically determined by repetitive tests. This basis matrix was used as an initial basis matrix to calculate the activation matrix

H of each $rmS_{xx}(\alpha, f)$ matrix. Because a cyclostationary component-related basis matrix should be obtained, the transposed $rmS_{xx}(\alpha, f)$ matrices were used in the process mentioned above, that is, the representative basis matrix contained inherent properties of cyclostationarity and we could extract a feature set consisting of the H matrices:

$$V \equiv WH \tag{4.4}$$

Because WH is an approximated matrix of V, the factors W and H were chosen using the minimization of the root mean squared (RMS) residual between V and WH. Through these procedures, $rmS_{xx}(\alpha, f)$ (54×889) of a long-term breathing sound was transformed into an NMF activation matrix (45×54). Based on the dimension reduced matrix, we calculated seven basic statistics, maximum, minimum, median, standard deviation, variance, kurtosis, and skewness, according to the rows and columns of H. As a result, these 693 (45×7 + 54×7) statistical values are the features of nonstationary analysis based on cyclostationarity and we defined this statistical data set as the second feature in this study. This feature can analyze the statistics of the activity of the cyclic spectrum magnitude based on the spectral or cycle frequency domain. Figure 4.3 shows the second feature extraction process.

4.3.6 Feature selection

For better classification performance, we conducted a feature selection process that eliminated redundant data in the features as mentioned above. We used the wrapper subset evaluation method, which is a flexible supervised attribute selector. We used a support vector machine (SVM) as a classifier of the evaluator and linear forward selection as an attribute search method. This process was performed using the WEKA framework [75], which embedded

Key scheme	Setting value	Parameter setting	
Attribute selection evaluator	WrapperSubsetEval	**Classifier**	SVM (default)
		Evaluation Measure	Accuracy (discrete), RMSE (numeric)
Attribute selection search	LinearForwardSelection	**Forward selection method**	Floating forward selection
		Perform ranking	True

Table 4.2 Parameter setting in WEKA for feature selection

various attribute selection methods.

Abovementioned linear forward selection technique can find smaller optimal attribute subsets from full attributes and can reduce a risk of overfitting. Finally it can provide higher classification accuracy [93]. This technique initially ranks all attributes and selects top-*k* ranked attributes by their scores that are obtained using a previous wrapper evaluator. Using a limited number of attributes and *m*-fold cross-validation, this search technique finds the optimal subset size. Therefore, a result subset has an exact size, and which is final feature set determined as the input of classifier.

The parameter setting in WEKA related to the feature selection used in this experiment is summarized in the following Table 4.2.

In Chapter 3, we also summarized the parameter setting values related to feature selection in WEKA as shown in the above table. Besides, we explain some of these parameters related to the feature selection method. Feature selection process is a search problem that selects a criterion function to validate a feature subset and then detect an optimal feature set based on this function. The criterion function based on the key scheme in the above table is a wrapper subset evaluator, and the search method is linear forward selection. The search process requires a computationally feasible procedure that avoids exhaustive searches. The search for an optimal feature subset can be divided into a bottom-up method (forward) and a top-down method (backward). The former is a way to add key features starting with an empty feature set, and the latter is a way to get the final feature set by removing the non-critical features from the complete measurement set. However, there is a nesting effect between the two methods, i.e. in the case of the forward method, the added feature cannot be deleted later, and in the case of the backward method, the deleted feature cannot be reselected. To solve this problem, various studies have been carried out. For example, floating forward selection is performed by Pudil et al.[94] (The selected parameter of the forward selection method in the table). Unlike conventional methods, this method allows flexible changes to approximate the result dimensions without pre-assigning the result dimensions when obtaining the best feature set. These values are not fixed and are 'float' so that the results of each step are not monotonously changed but are floating. Although this method does not always provide the best subset features, it is known to be better than other search methods and is more efficient than the branch & bound method, which is known for its breakthrough in feature subset search.

4.4 Evaluation

Using the searched feature subset, we performed three classification tests. For

Cross-validation scheme		Leave-one-out	
Classifier	**Key scheme**	**Option**	
SVM	Build LogisticModels	False	
	Complexity	1	
	Epsilon	1.0E-12	
	Filtertype	Normalize training data	
BayesNet	Estimator	SimpleEstimator	
		Alpha	0.5
	Search algorithm	K2	
		Score type	Bayes
Logistic	ridge	1.0E-8	
	maxlts	-1 (unlimited)	
Random Forest	maxDepth	0 (unlimited)	
	numFeatures	0 (unlimited)	
	numTrees	110	
Refer to the table in Chapter 3 for a detailed description of each parameter			

Table 4.3 Evaluation test related setting parameter

two types of features, transition probability and cyclostationary based information, we performed the individual classification test. Then we conducted the final classification test with complete feature subset to validate the performance of multiple OSA severity classification. All the classification tests provided the accuracy of the four OSA severity classifications using leave-one-out cross-validation (LOOCV). Additionally, we calculated the sensitivity and specificity based on the multiple OSA severity classification results to compare classification performance. In this section, all tasks were conducted using the WEKA framework [75], which was employed in previous processes and provides various feature selection methods and machine learning classifiers. In this experiment, evaluation was performed using random forest, Bayes network, logistics, and SVM as in the case of the Chapter 3 experiments, and a classifier showing optimal results was selected by comparing the results. The final selected evaluation test related setting parameters are shown in the following Table 4.3. Each parameter value was selected through a repetitive heuristic test or a grid search.

4.5 Results

4.5.1 Feature subsets for OSA severity classification

Using the feature selection method, we obtained 18 features from two types of features. Three were selected from the temporal analysis features, and the remainder were selected from the nonstationary features. Table 4.4 shows the complete feature list of the feature subset. In this table, we show the results for the rank, base, observation, statistics, and sequence number. The "Rank" was determined using an attribute search method: wrapper subset evaluation feature selection method. A higher rank means that the associated feature is more significant for the classification task. In this study, the final nonstationary analysis feature representation, called an NMF activation matrix, has a 45×54 dimension. Its x-axis represents the spectral domain, and the y-axis represents the dimension-reduced cycle frequency α domain. Based on this matrix, seven basic statistical values were calculated according to each axis: α and f index. The "Base" indicates a base axis for observing the statistical activation status of the other axis, which is presented in the "Observation" column. The "Statistics" represents the type of statistics that were calculated for the observation axis. The "Sequence number" represents an index of a particular base axis. For example, if [base = α, observation = f, statistics = maximum, sequence number = 40], then we select the maximum of the f

Cyclostationary analysis based feature subset				
Rank	**Base**	**Observation**	**Statistics**	**Sequence number**
1	α	f	maximum	40
2	α	f	maximum	42
3	α	f	variance	34
4	α	f	kurtosis	8
5	f	α	kurtosis	42
6	f	α	maximum	24
7	f	α	standard deviation	42
8	f	α	variance	24
9	f	α	median	24
10	f	α	median	45
11	f	α	median	46
12	f	α	mean	7
13	f	α	kurtosis	1
14	f	α	kurtosis	2
15	f	α	skewness	2
Temporal analysis subset features				
Rank	Energy level transition information			
16	(1×1) from Level 1 to Level 1			
17	(3×4) from Level 3 to Level 4			
18	(4×1) from Level 4 to Level 1			

Table 4.4 Final feature subset selected from the original feature set

index domain's activation statuses associated with the 40th index of the dimension-reduced cycle frequency index domain as a feature. An example of this analysis procedure is illustrated in Figure 4.4.

Various statistical values were selected from the original feature representation. Because the statistical results were calculated from one column or row of a matrix, and the distribution of activations was important, the subset features

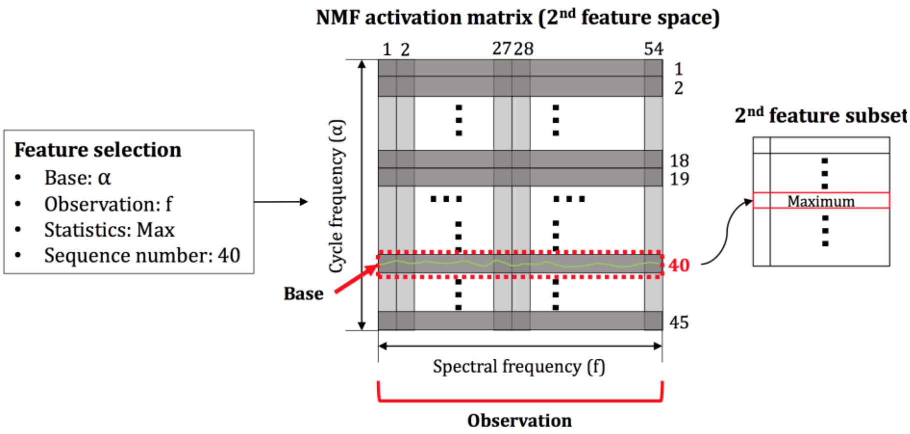

Figure 4.4 Feature selection from the NMF activation matrix.

x-axis: the spectral domain, y-axis: the dimension-reduced cycle frequency (α) domain

included many statistical descriptors of the shape of distribution, such as a kurtosis or skewness. Figure 4.5 shows four NMF activation matrices averaged based on these 15 nonstationary subset features. To observe the overall distribution of the magnitudes corresponding to each dimension-reduced cycle frequency α and spectral f index pair, we calculated the average matrices, including the 15 nonstationary subset features of all subjects, and adapted a Gaussian filter to check the distribution of each matrix.

Final feature subset was selected from the original feature set. From the original feature set that consisted of temporal analysis and cyclostationary analysis based features, a dimension reduction technique and a feature selection method were applied for obtaining more optimized feature subset and better classification accuracy.

According to Figure 4.5, the four average NMF activation matrices have different distributions on OSA severity. The normal group demonstrates more widespread spectral activations associated with a dimension-reduced cycle frequency index domain than the others. The mild group’s distribution is

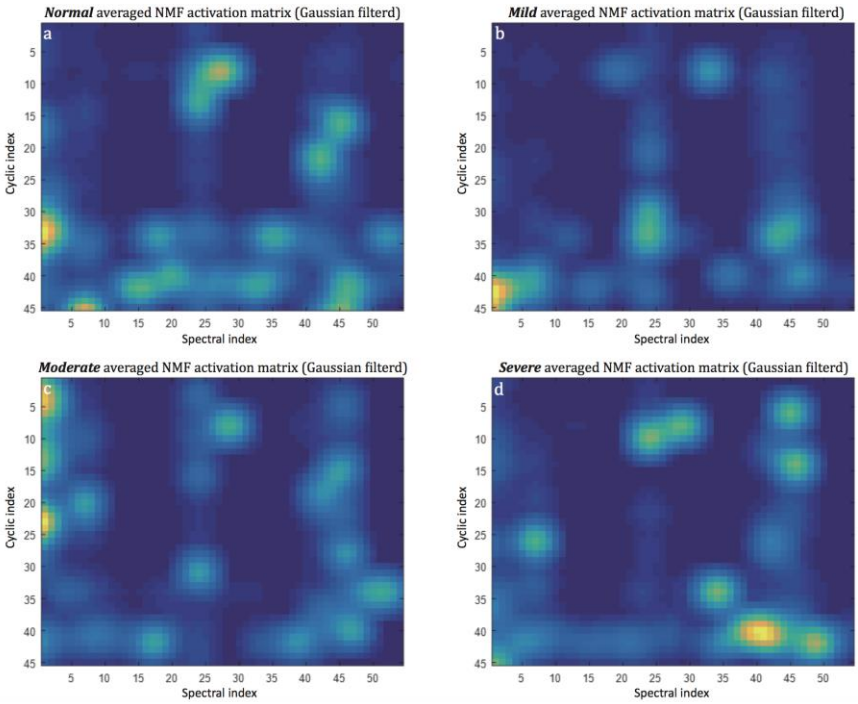

Figure 4.5 Averaged four NMF activation matrices based on the final cyclostationary analysis based dimension reduced feature subset. The NMF activation matrices show different distributions of corresponding magnitudes for each dimension-reduced cycle frequency α and spectral f index pairs

relatively sparser than any other groups, in particular, it has no evident cyclic activations related to a low range of the spectral index.

The activation distribution of the moderate group demonstrates relatively high cyclic activation at a low range of the spectral index. The severe group's distribution demonstrates higher spectral activations at a high range of cyclic indices and higher cyclic activations at a high range of spectral indices when compared with the other groups. Moreover, the significant highest activation regions of the NMF activation matrices are all different according to the OSA severity.

Based on these observations, we can assume that our cyclostationary analysis

Temporal analysis feature		1×1	3×4	4×1
ANOVA		***	***	***
Tukey HSD	moderate-mild	*n/s*	*n/s*	*n/s*
	normal-mild	**	*n/s*	*n/s*
	severe-mild	***	***	***
	normal-moderate	***	*n/s*	**
	severe-moderate	**	***	***
	severe-normal	***	***	***

** (0.01 < p < 0.05), ** (0.001 < p < 0.01), *** (p < 0.001), n/s: not significant*

Table 4.5 Analysis results of the temporal feature subset

based feature set is useful for classifying the breathing sounds into four OSA severity groups. Regarding temporal analysis subset features, three features were selected from the original set and they represented the transition probability of the approximated breathing sound's energy values. The selected transition information indicated that the silent section and those associated with a predefined OSA candidate were important for OSA severity classification.

By performing statistical analysis using these temporal analysis feature subset, we were able to identify the differences between the four OSA severity classes. Using analysis of variance (ANOVA) and Tukey's honest significance difference (HSD) test, we verified that all three temporal analysis features were significant ($p < 0.05$), and most class-pairs (normal-mild, normal-severe, normal-moderate, mild-severe, and moderate-severe), with the exception of the moderate-mild class-pair, demonstrated a significant difference ($p < 0.05$) regarding these features.

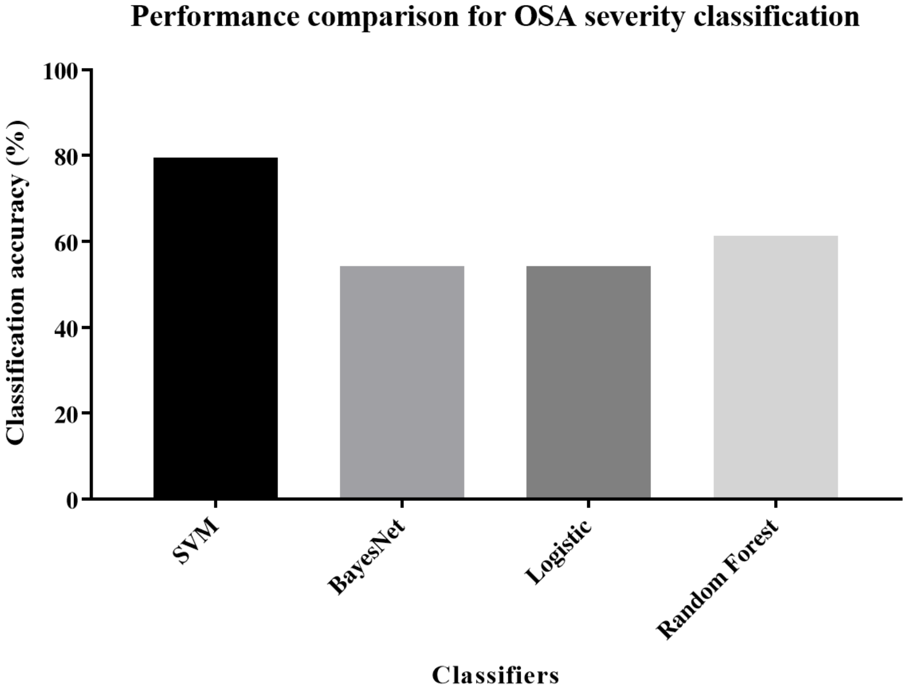

Figure 4.6 Performance comparison for OSA severity classification

Group (OSA)	True positive rate	False positive rate	Precision	Recall	F-measure	ROC Area	PRC Area
Normal	0.75	0.02	0.94	0.75	0.83	0.93	0.81
Mild	0.86	0.16	0.64	0.86	0.74	0.81	0.59
Moderate	0.67	0.08	0.74	0.67	0.70	0.77	0.58
Severe	0.91	0.02	0.95	0.91	0.93	0.98	0.92
Weighted average	0.80	0.07	0.82	0.80	0.80	0.87	0.72

Table 4.6 Detailed results of cross-validation (using SVM)

4.5.2 Classification Test with the Subset Features

Using the subset features, we performed the four-OSA severity classification

OSA \ Classified as	Normal	Mild OSA	Moderate OSA	Severe OSA
Normal	15	4	1	0
Mild	1	18	2	0
Moderate	0	6	14	1
Severe	0	0	2	19

Table 4.7 Four-OSA severity classification result with leave-one-out cross-validation

test. We trained the classification model using an SVM [95-97] with a linear kernel and confirmed the model's performance using LOOCV. All the experiments were conducted using a WEKA-implemented classifier and validation tools, and their configuration settings initialized with default values (Exclude tree number in the random forest). Also, to select the best classifier, classification experiments were performed using various classifiers built in WEKA framework such as random forest, Bayes network, logistics and the SVM was finally selected among them.

The detailed results of cross-validation using SVM are shown in Table 4.6. The moderate OSA group has the lowest true positive (TP) rate and AUC, while the severe group has the highest TP rate and AUC.

The classification accuracy of the four-OSA severity classification test was 79.52 %. Table 4.7 shows the classification result as a confusion matrix. In the moderate OSA group, it showed classification errors on the mild OSA group. Moreover, the majority of the normal subject group's classification error is related to the mild OSA group. The weighted-average AUC for the four-class classification test was over 0.8, and it means relatively good performance [77]. Using Table 4.7, we can also obtain the binary classification result to classify normal subjects (AHI $<$ 5) and OSA patients (AHI $\geq$ 5). The binary classification results show that the sensitivity is 98.0 %, specificity is 75.0 %,

Method	Participants	Microphone location	OSA groups	Sensitivity	Specificity
				Accuracy (%)	
Nakano (2004) [62]	383	Neck (Contact)	Two	93	67
Abeyrantne (2005) [45]	16	Patient vicinity (40~70 cm)	Two	100	50
Azarbarzin (2013) [64]	57	Neck (Contact)	Two	92.9	100
			Four	77.2	
Behar (2015) [63]	856	Face (Contact)	Two	69.5	83.7
Proposed	83	Patient vicinity (170 cm)	Two	98.0	75.0
			Four	79.52	

Table 4.8 Method comparison between related studies using snoring sounds

and the classification accuracy is 92.78 %.

Referring to the comparison proposed in the related study [64], in Table 4.8, we compared our method with other studies regarding the number of subjects, microphone's location, number of OSA groups, and performance.

Since there is no standardized performance comparison framework for studies using sleep breathing sounds [64], our study may not be evaluated as being objectively superior. However, this table is presented to check our research performance level against the previous related studies.

4.6 Discussion

In this study, we demonstrated that our new sleep breathing sound analysis method can provide relatively high performance for multiple OSA severity classification. We hypothesized that the energy transitions related to the general breathing sounds, snores, and silence in a time series, and the cyclostationarity-based nonstationary characteristics of the sounds associated with obstruction and vibration in the upper airway could be used as significant features that represent long-term breathing sounds of a participant. This hypothesis has been validated using experiments that classify sleep breathing sounds into OSA severity classes based on the AHI.

In this study, we extracted cyclostationary analysis based features using an entirely new approach. We calculated an average CS from a subject's sleep breathing sounds, for which the average time was over four hours. Then the NMF method was adapted for drastic dimension reduction, and a specific feature selection method was also applied to search significant feature subset. To our knowledge, this is the first study using cyclostationary analysis based information with NMF and feature selection for sleep breathing sound analysis. The results of this approach show that cyclostationary activation, which represents a hidden periodicity of data related to particular spectral bands, could be a special feature for sleep breathing sounds. Using this method, we summarized the nocturnal breathing sounds of each subject, with particular properties that were associated with the spectral characteristics and hidden periodicity of the sounds. We verified that this feature represented significant differences between breathing sounds, which were grouped according to the OSA severity class, as shown in Figure 4.5. For the normal group, the NMF activation matrix revealed that a wide spectral band area was associated with the narrow high indexed cycle frequency band area. By contrast, the moderate and severe OSA groups presented different characteristics. For these groups, a wide cycle frequency band area associated with a particularly high indexed spectral area was activated in the matrix. We found that these properties reflected the special spectral characteristics of representative breathing sounds of each subject's nocturnal breathing sounds.

The temporal analysis features were the transition probability of breathing sound energy in time series. We adapted basic OSA detection criteria that are related to the silent interval between snoring sounds, for example, apneic events greater than 10 s in duration [86]. In Table 4.4, the final temporal analysis subset feature consisted of three types of transition information: (1×1) – silence level, (3×4) – energy transition from the high energy level to the OSA candidate level, and (4×1) – energy transition from the OSA candidate level to the silence. We showed that all searched features were statistically significant in Table 4.5 in which the appearance rates of purely silent sections and the sections of energy transition to silence could be an important feature for the OSA severity classification task. The aforementioned two feature subset could be influenced by the sound recording quality concerning the recording performance or location of the microphone. Although the microphones are located on the ceiling and are experimented in a special place called the PSG room, there is a limit to applying the developed analysis framework directly to the general user environments. However, unlike previous studies, it is clear that the recording quality of the sound used in this study is not comparatively good because the professional recording environment including special microphones is not considered in this study. Therefore, attempting to distinguish OSA severity using only sounds in such a situation can be a clear distinction from other related studies. The performance and location of the recording device are an important consideration when applied to the actual user environment. As shown in Section 4.3.1, personal portable smart devices can embed better performance microphones than that used in this experiment. Thus, if additional algorithm tuning or alteration is performed due to changes in the experimental parameters, such as changes in sound quality due to the use of better performance microphones and proximity to the user, we expect that proposed analysis framework will provide better results in tasks related to sleep breathing sounds. In Table 4.8, we compared our study with previous related research. Unlike previous studies, our proposed method did not use a special body contact-type microphone and did not perform any snoring detection process with breathing sounds. Furthermore, the sleep breathing sound utilized in this experiment was recorded via the microphone installed farthest from the patient compared to other studies. We only divided the sleep breathing sound into predefined window lengths and generated the feature representation and subset using the features mentioned above from all the windows. We intend to use this method in a screening system which can provide the personal OSA critical alarms or can provide notification of whether PSG test is necessary.

4.7 Summary

In this study, we proposed an OSA severity classification method for a preliminary PSG test using particular features of nocturnal sleep breathing sounds. Unlike previous research, this study did not use any current known features used in the field of sound analysis. Instead, the energy transition probability information of the audio signal in the time domain, and the cyclostationarity-based nonstationary characteristics in the spectral and cycle frequency domains were used in this experiment. Using these two features, the proposed method showed the very competitive classification performance of 79.52 % accuracy, 87.00 % average AUC for the multiple OSA severity classification, and also 92.78 % for OSA patient detection test. These results indicate that a proposed method could be a promising approach to identify the multiple OSA severity of suspected patients and provide proper information to individuals for a preliminary PSG screening test. The limitations of conventional PSG, such as the high cost, inconvenience, complex measurement method, and Inaccurate results due to sleep variability, lead to an increased demand for a preliminary PSG screening test in various environments, such as the home. In the proposed method, the sounds were recorded only in the clinical test rooms, so it can be considered that the experiment did not consider the actual environments. However, since the general private bedroom environment is not much different from the PSG room, and the sounds were not recorded under particular controlled conditions, the proposed algorithm is expected to perform reasonably well in the universal circumstances by parameter tuning and supplementing algorithms based on various experimental environments in the future.

The proposed method also has some limitations. For more practical applications, there is a need to apply various noise reduction and cancellation techniques to the acquired sounds or framework of research. Additionally, experiments with many more patients should be conducted to make our method more robust and reliable. Furthermore, to obtain more accurate classification performance, additional algorithms can be considered in the preprocessing or feature extraction process. In particular, the properties related to cyclostationarity can be expected as a good input data of feature learning techniques for a deep neural network; thus, we will consider this in a future study.

The present study will contribute to the development of screening technology for a specific medical inspection using limited data, and we expect that this technique will be applied to various healthcare service platforms to supplement a preliminary examination of sleep disorders.

Chapter 5. Patient OSA Severity Prediction using Deep Learning Techniques

In this chapter, we applied deep learning techniques to the previous studies. For the SRBD-related snoring classification performed in Chapter 3, we conducted the feature learning of CS using convolutional neural networks. With this learned model, we also performed the patient's OSA severity classification task based on partial sound section extracted from patient's long-term sleep breathing sound.

5.1 Introduction

Recently, as we have seen through Google Alpha-Go*, deep learning is being used in a variety of studies to learn features from certain data and to create classification models. The artificial neural network that underpins deep running has had two boom-ups from the 1940s to the present. The back-propagation method was devised at the time of the second boom-up, and the research field seemed to make a great leap forward. However, it was not properly learned when the number of layers of the neural network was increased, and it was evaluated that the performance was lower than that of the machine learning technique because it was impossible to explain how to set various parameters (number of layers, unit) optimally. However, as Hinton et al. [98] announced a

* Google AlphaGo: https://deepmind.com/research/alphago/

deep belief network (DBN), the neural network research turned out to be a turning point. DBN is a graph model with a multi-layer structure similar to a general neural network. They propose a learning method based on pre-training using a restricted Boltzmann machine (RBM) and a greedy algorithm to limit learning difficulties in the course of increasing number of layers. After that, the validity of multi - layer neural network for speech recognition and image recognition has been verified through various studies, and the classification performance results of these studies have overwhelmed the existing ones. Currently, there is active research on performance improvement through deep learning in various fields. These studies use three representative methodologies such as fully-connected neural networks, convolutional neural networks (CNN), and recurrent neural networks (RNN), according to the data. Fully-connected neural network or RNN is used for speech recognition, and natural language processing, and also CNN is applied for image recognition tasks.

CNN is based on neuroscientific facts about the visual cortex in the brain of an organism and models simple and complex cells in the retina that selectively respond to specific input patterns by light sources [99]. Simple cells usually have strict regioselectivity, but complex cells do not. In a typical CNN structure, the convolutional layer and the pooling layer are the main components. The composite layer is a simple cell, and the pooling layer placed after a composite layer is a model of complex cells. Therefore, the pulling layer lowers the position sensitivity of features extracted from the composite product layer so that the output of the feature does not change even if the position of the feature changes slightly in the image. The features of this convolutional layer are equal to the weights of the layer. In the case of CNN, we optimize the weights of the composite products by the gradient descent method, similar to the general feedforward neural network. Therefore, the weight of the optimized composite

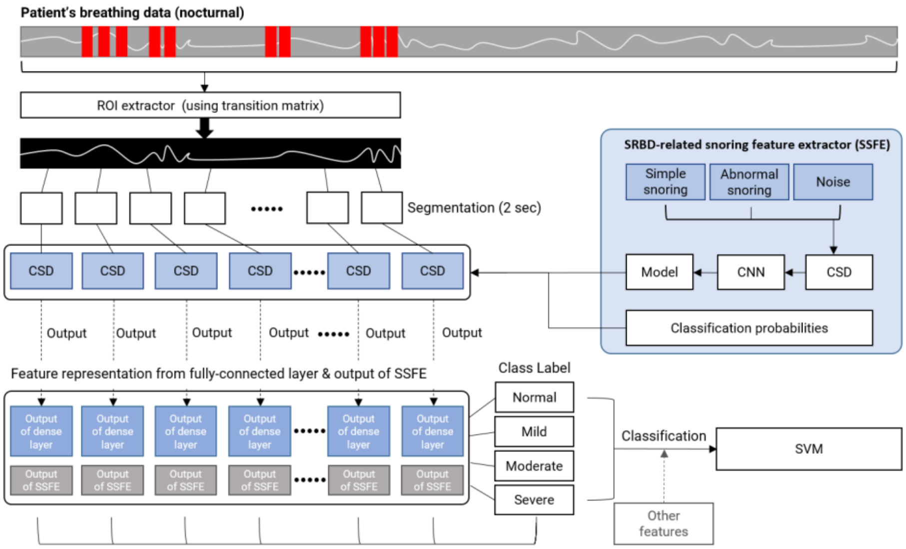

Figure 5.1 New overall architecture of the SRBD-related snoring and OSA severity classification

product layer is an optimization filter itself.

5.2 Methods

In this chapter, we applied deep learning techniques in the study of sleep disorders using sleep breathing sounds. As far as we know, there has not yet been any application of deep learning in this sleep breathing analysis fields. In this chapter, we performed the SRBD-related snoring carried out in the previous Chapter 3 using CNN. Then, the learned model was applied to the sleep breathing data set to obtain the classification probability value for the window of the corresponding signal.

Furthermore, we consider the outputs of fully-connected layers of CNN as the final feature set and the OSA severity classification was performed as in previous Chapter 4. From the CNN structure used in this experiment, we extracted two feature representations, the first is the processed output of a specific fully connected layer and the second is the probability of 3 snoring classes derived from the final output layer which is applied a softmax. The subject's sleep breathing sound data set was the same as in Chapter 4. However, a 10-minute partial breathing sound, which was extracted separately without using the entire sleep breathing sound, was utilized in this experiment.

Through the experiments in this chapter, we tried to compare the classification performance of traditional hand-crafted features with features derived from feature learning process using deep learning techniques. Particularly, we tried to maximize the effectiveness of the proposed framework in practical applications by using partial sleep breathing sounds. Figure 5.1 shows the new overall architecture of the SRBD-related snoring and OSA severity classification system proposed in this chapter. Each key process will be described in the following sub-chapters.

This experiment was performed using Keras [100], a high-level neural network API which is written in Python, on Ubuntu 14.04 LTS. It also runs on top of Theano*, which is a Python library for deep learning tasks, using a GPU (Geforce GTX 980Ti).

5.2.1 Feature learning of SRBD-related breathing sounds

First of all, we confirmed the feature learning possibility using deep learning techniques, and the practical classification performance using the CS, which is a symbolic feature representation of sleep breathing sounds used in previous experiments [44, 61]. In this study, we extracted simple snoring, hypopnea-related snoring, and apnea-relate snoring based on the recorded sleep breathing

---

* Theano 0.9 (http://deeplearning.net/software/theano/)

sounds during PSG test, as in the case of Chapter 3's research processes. Also, the silent section or noise region was defined as no-event, and it has been extracted from the sleep breathing sounds.

Unlike previous experiment [61], hypopnea-related snoring and apnea-related snoring were integrated into abnormal snoring and generated new data set consisted of three types of events: simple snoring, abnormal snoring, and no-event. In this study, we aimed to obtain a better performance SRBD-related snoring sound classifier using deep learning technique and to provide OSA occurrence result of the patient with simple way. Therefore, we judged that this breathing event scheme could better reflect the actual situation.

Each event consists of 200 sound data of 2 seconds length. Therefore, the total number of event data used in the experiment is 600. The sound data of 2 seconds length is converted into CS [44] through the cyclostationary analysis introduced in Chapters 2 and 3. In previous experiments, complex processes such as dimension reduction and statistical analysis applied to this CS for extracting hand-crafted features. However, in this experiment, we only applied feature learning process using the CNN on a CS. The CNN used in the experiment has a typical structure consisting of convolutional layers, pooling layers, and fully-connected layers. The proposed CNN structure is summarized in Table 5.1.

The CS image input to the CNN structure is an image file stored in the system by down-sampling the CS matrix calculated using cyclostationary analysis to 174×132 pixels. These image files were created as a dataset file in one Hierarchical Data Format (HDF) [101] linked to individual snoring labels. The CNN used in the experiment was comprised of repeated three convolutional layers and pooling layers and had four fully-connected layers, all layers used dropout procedure for improving generalization error, and each fully-connected

idx	Type	Output shape	Param #	Kernel (Pool size)/ stride
0	input	(132, 174, 1)		
1	Conv2D	(132, 174, 32)	320	3x3 / 1
	tanh	(132, 174, 32)		
	MaxPool	(66, 87, 32)		2x2 / 1
2	Conv2D	(66, 87, 64)	18496	3x3 / 1
	tanh	(66, 87, 64)		
	MaxPool	(33, 43, 64)		2x2 / 1
3	Conv2D	(33, 43, 64)	36928	3x3 / 1
	tanh	(33, 43, 64)		
	MaxPool	(16, 21, 64)		2x2 / 1
4	Dense	2048	44042240	
	BatchNorm	2048	8192	
	ReLU	2048		
5	Dense	1024	2098176	
	BatchNorm	1024	4096	
	ReLU	1024		
6	Dense	512	524800	
	BatchNorm	512	2048	
	ReLU	512		
7	Dense	3	1539	
	Softmax	3		

Table 5.1 Detailed description of the best CNN of data set

layer performed batch normalization before activation.

5.2.2 Partial sleep breathing sound extraction

Experiments to classify the patients' OSA severity in Chapter 4 used a very long breathing sounds, excluding sound sections of some sleep stages, from the total

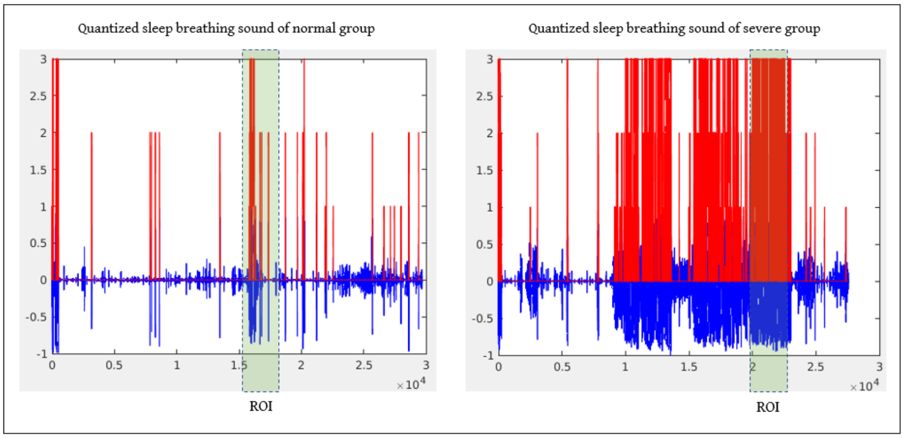

Figure 5.2 ROI selection from sleep breathing sounds

sleep breathing sounds. This approach is a method of extracting and analyzing features from most recorded sleep breathing sounds from one patient and confirming the results, but it is certainly not efficient. Therefore, the previous analysis system should record most of the patient's sleep breathing sounds and save with a particular file format in storage, and also should use the entire stored sound file in the analysis process.

In Chapter 4, we have investigated the transition matrix [44] that reflects the characteristics of the sleep breathing sounds in the time domain. To transform the sleep breathing sound of one patient into this matrix, we quantified the raw sleep breathing sounds of the patient to several quantized levels according to the labeling criteria. That is, the sleeping breathing sounds of one patient were labeled with silence interval, low-energy sound interval, high-energy sound interval, and apnea suspicion interval (intervals longer than 10 seconds between low or high-energy sound interval) [86]. In this experiment, the region of interest (ROI) was extracted from the total sleep breathing sounds of one patient using this label information.

In the same way, as in the previous study, the sleeping breathing sounds were converted to energy signals and based on this, and the moving average was sequentially calculated to smooth the signal. Finally, the signal was quantized according to a predetermined criterion. The number of occurrences of a high-energy interval (label = 3) and apnea suspect interval (label = 4) was calculated for each specific length (= 10 minutes) in the entire signal quantized with four labels. Finally, the section of the specific length with the greatest number of label occurrences was designated as the ROI for the sleep breathing sounds of the individual patients. An example of this process is shown in Figure 5.2. In this process, we assumed that the OSA severity of the patient could be represented by a region which includes large snoring sounds and OSA suspicion sections frequently occurs in the time domain. Of course, it is not easy to distinguish multiple OSA severities with this information alone. Since OSA severity cannot be assessed by the loudness or frequency of occurrence of the patient's snoring, and furthermore, OSA suspicion sections are not always real OSA intervals. Therefore, we applied the previous SRBD-related breathing sound classifier as

a feature extractor on the partial sleep breathing sounds. A pseudo-code for partial sleep breathing sound extraction so far is given in Algorithm 5.1.

Algorithm 5.1 Pseudo-code for ROI extraction
input Output of step 9 of the Algorithm 4.1 (energy signal smoothing using moving average)
1: ***SumE*** = Sum of label 3, 4 occurrences in each 10-minutes window
2: ***ROI*** = if ***SumE*** > Total number of predetermined label occurrences
output ***ROI***

5.2.3 OSA severity classification using partial breathing sounds

The SRBD-related breathing sound classifier outputs three classification probabilities for the CS calculated from the breathing sounds at a specific time in the final fully-connected layer. As shown in Table 5.1, these probabilities are generated by the last fully-connected layer. In this experiment, the result values processed by a specific fully-connected layer, which is an output of the sixth layer in Table 5.1, and probability values of output layer are used as feature representations. In fact, fully- connected layers, which is a component of the general neural network, are utilized in the process of actually classifying the input data using the learned features in the previous convolutional layer.

We define this set of layer's outputs, derived sequentially from the window of the signal, as a specific feature representation in a task that classifies OSA severity using partial sleep breathing sounds. In this experiment, the length of the partial sleep breathing sounds is set to 10 minutes, and sixth layer's and final layer's output through the SRBD-related breathing sound classifier is generated for each 2-seconds window. Therefore, 3×300 and 512×300 formed feature representations were generated per patient data. We transformed it into 1×153,600 and 1×900 forms respectively by flattening them and applied feature selection using logistic regression. Similar to the feature selection procedure in Chapter 4, we used a logistic regression in this study as a base classifier to evaluate a subset of features [102]. We searched for feature subsets with various sizes and selected the optimal subset using 5-fold cross-validation. For the first feature representation, a subset of 10,000 features was finally selected and its validation result (mean % and 95% CI) was 0.95(±0.09). Also, for the second feature representation, a subset of 138 features was selected, and its validation result was 0.86 (± 0.11).

Validation scheme	10-repeated hold out test	
Classifier	SVM (Python scikit-learn)	
	Parameters	**Option**
	Kernel	Linear
	Complexity	0.01 or 1
	Tolerance	1E-3
	Iterations within solver	No limit
	Random state (seed)	1,3,7,14,28,35,42,55,97,100

Table 5.2 Parameter settings for validation test

Once the feature subsets have been determined, we performed repeated hold-out tests using existing dataset. Ten random datasets were created by setting 33 % of the dataset as the test set and the remains as the training set. The classification model was generated using SVM [95], and actual ten classification results were calculated. The various parameter settings used in this classification experiment are summarized in Table 5.2. Besides, we confirmed the classification results by combining the feature representation created through the feature learning of this experiment and the hand-crafted feature proposed in Chapter 4. Through these experiments, we confirmed the effect of the combination of specific features on the improvement of classification performance. In other words, we focused on identifying whether the hand-crafted feature could be used as a complement to the learned feature. A pseudo-code for OSA severity classification using CNN and partial sleep breathing sounds so far is given in Algorithm 5.2.

Algorithm 5.2 Pseudo-code for OSA severity classification using CNN
input CS result of each signal window ***wCS***, SRBD-related snoring classification model ***M***, Evaluation function for feature selection ***FS*** = logistic regression, Full dataset ***D***
1: Feature representations (***FR***) = ***M(wCS)***
2: Feature subset (***fSubset***) = ***RFE(FS)***
3: Split the training (***TR***, 67%) and test set (***TE***, 33 %) from ***D***
4: Training subset: ***subTR = fSubset(TR)***, Test subset: ***subTE = fSubset(TE)***
5: OSA classification model 1 (***mOSA1***) = ***SVM(subTR)***
6: Concatenate feature subset (***cSubset***) = ***[subTR, Other features]***
7: OSA classification model 2 ***(mOSA2) = SVM(cSubset)***
8: Holdout test: ***Accuracy*** = ***mOSA1(subTE) and mOSA2([subTE, Other features])***
output ***Accuracy, sensitivity, specificity***
RFE: Feature ranking with recursive feature elimination *Other features: Hand crafted features of previous chapter*

5.3 Results

5.3.1 OSA severity classification using output of dense layer

In the SRBD-related breathing sound classification experiment, the CS was calculated using the cyclostationary analysis for individual respiratory sound events, and the classification model was created by applying the CNN structure proposed in the previous section and confirmed the results.

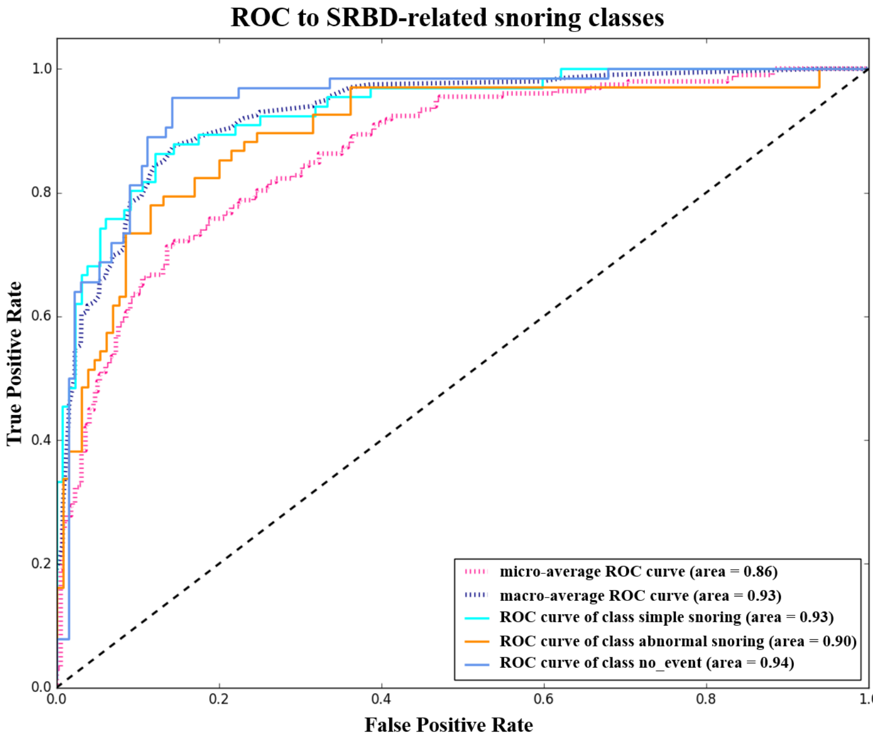

Figure 5.3 ROC to SRBD-related snoring classes

We confirmed 82.83 (± 4.78) % classification performance in 10-fold cross-validation for simple snoring, abnormal snoring, and no-event data. Figure 5.3 showed that the average AUC of the three-class classification test was 0.8 or more, and it indicated that the performance is good in a clinical practice [77]. The classification model developed through this experiment is an event classifier that can distinguish between normal-snoring and abnormal-snoring for breathing sound during sleep. We applied this classifier to the partial sleep breathing sounds extracted from the whole sleep breathing sounds of one patient. Since the classification model was not created using a large number of patients' snoring data, this classifier's performance cannot be guaranteed from the system when applied to partial sleep breath sounds of other patients. Therefore, we did not directly use the classification results of this classifier in

OSA severity classification. Instead, the two outputs from the events classifier were considered as feature representations in the OSA severity classification task.

5.3.2 OSA severity classification using output of dense layer

Sixth layer output of the CNN structure was considered as feature representation in the multiple or binary OSA classification task. The output of a fully-connected layer is a value derived from processing the features extracted from the convolutional layer in the CNN structure for classification as in the layers of a typical neural network. Based on the series of fully-connected layers combined after the convolutional layers, the output of the third layer was selected as the feature. Therefore, the selected feature representation in this experiment has the activation information of the features derived from the convolutional layer by the following equation.

$$output = activation(\,(input \cdot kernel) + bias\,) \tag{5.1}$$

where, activation function is ReLU, the kernel is a weight matrix created at each layer, and bias is a bias vector generated at each layer.

The $1 \times 153{,}600$ formed feature representation was extracted from the 10 minutes partial breathing sounds. It was generated by concatenating 300-row vectors with 512 elements. Then, feature selection was performed through logistic regression and, finally, 10,000 features were selected through repetitive experiments. The classification model was generated by the SVM (c=0.01) of the linear kernel, and the classification performance of 0.85 (± 0.05) was confirmed in multiple OSA Severity classification tasks through 10-repeated hold-out tests. In a binary classification test to determine whether a particular patient had OSA symptoms, the average sensitivity and specificity were 90.73

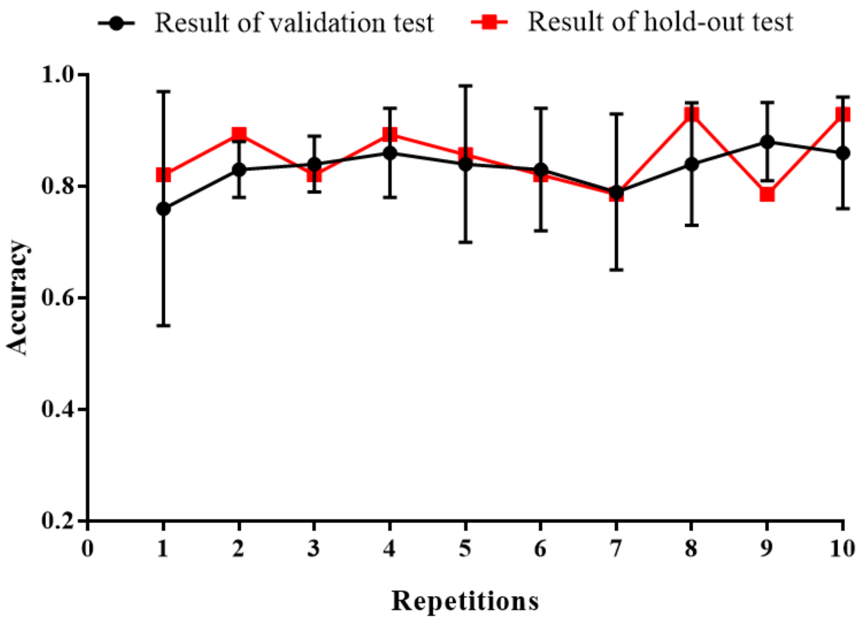

Figure 5.4 Results of multiple OSA severity classification using output of dense layer

(±5.65) % and 95.25 (±6.17) %, respectively, by the AHI value of 5 and 10-repeated hold-out tests. Figure 5.4 shows the individual 5-fold CV results in the repeated hold-out test and the classification results in the separate test sets.

5.3.3 OSA severity classification using output layer

We also considered the classification probability information values for the three classes derived from the outputs of event classifier as another feature representation in the OSA severity classification task. In fact, this probability value indicates whether any breathing signal belongs to three events, simple snoring, abnormal snoring, or no-event. However, if we do not associate event labels with this, this set of probability values can be regarded as a characteristic value of a signal which is derived from a CNN. Therefore, the event classifier derived from the classification of respiratory sounds is used as a feature extractor of the signal in the OSA severity classification.

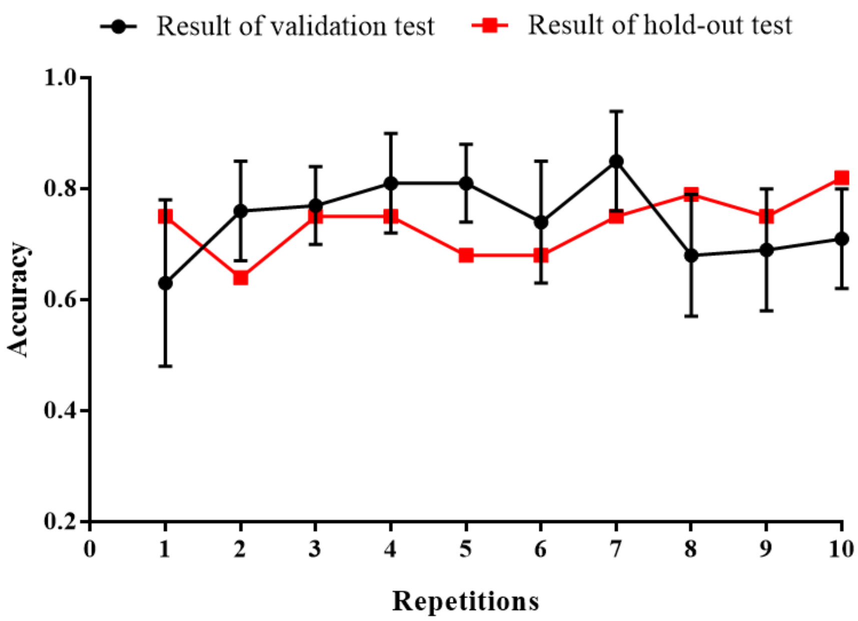

Figure 5.5 Results of multiple OSA severity classification using output layer

The 1×900 formed feature representation was extracted from the 10 minutes partial breathing sounds using preceding feature extractor. Feature selection was performed through logistic regression. Finally, 138 features were selected through repetitive experiments. The classification model was generated by the SVM (c=1.0) of the linear kernel, and the classification performance of 0.74 (± 0.05) was confirmed in multiple OSA Severity classification tasks through 10-repeated hold-out tests. In a binary classification test to determine whether a particular patient had OSA symptoms, the average sensitivity and specificity were 83.29 (±9.73) % and 84.75 (±15.16) %, respectively, by the AHI value of five. Figure 5.5 shows the individual 5-fold CV results in the repeated hold-out test and the classification results in the separate test sets.

Dataset	Method	Target for feature	Multiple OSA severity	OSA detection (%)	
				Sensitivity	specificity
ROI	Feature learning	DL	0.85	90.73	95.25
		OL	0.74	83.29	84.75
Compound	Feature learning + Hand-crafted	OL + TP	0.77	86.03 (±10.35)	85.25 (±14.36)
Full data	Hand-crafted	TP	0.63	87.3	65
		CS + TP	0.79	98	75

DL: Output of CNN's dense (Fully-connected) layer

OL: Classification probability from CNN's output layer

TP: Energy signal transition probability

Table 5.3 Results of new OSA severity classification / OSA detection

5.3.4 OSA severity classification using compound feature subset

Also, we compounded transition-matrix-based features that reflect the temporal characteristics of patient sleep breathing sounds obtained in Section 4 with the above feature subsets. Experimental results showed that the synthesis of temporal features derived from partial sleep breathing sounds did not affect the classification results. From this, it could be judged that the probability value is significant only if the signal transition probability is obtained from the respiration of the patient for a considerable long term. However, the temporal features of the long-term sleep breathing sound slightly increased the classification results. This is the case when compounding the feature subset of the CNN's output layer, and transition probability obtained in Chapter 4. In the multiple OSA severity classification experiment, the compounded feature subset showed 0.77 (± 0.07) classification accuracy in a 10-repeated hold-out test. In the binary classification test, the average sensitivity was 86.03 (±10.35) % and the specificity was 85.25 (±14.36) % by the AHI value. Table 5.3 summarizes the classification results of these experiments. This table also included the results of the previous Chapter 4 which corresponds to full data in the dataset category in the table. We cannot compare the results directly because the methodology of the experiments is not completely consistent, but we can confirm the advantage of CNN-based feature learning method.

5.4 Discussion

In this experiment, we performed feature learning through CNN in the CS calculated by the sleep breathing sounds and created a classification model that conforms to the individual task. In fact, the essence of this experiment is the improvement of the SRBD-related snoring sound classification discussed in Chapter 3 using the latest deep-learning techniques. This is because it can be utilized as a more accurate event detector and enables better characteristics extraction from sleep breathing sounds. Therefore, in this experiment, more snoring samples than the previous experiment were used for generating classification model, and the CS was calculated from each sample using the cyclostationary analysis employed in Chapters 3 and 4. We transformed the calculated CS into a two-dimensional image and applied it to the CNN structure consisting of three convolutional layers and three fully-connected layers to perform feature learning and classification. The learned features in the classification experiment of SRBD-related breathing sound showed higher classification performance than the classification model proposed in Chapter 3.

In the experiment in Chapter 3, we calculated CS from a snoring sound sample and obtained the covariance between the CS magnitude values that match one spectral frequency value and calculated the summation of the subregions of the image by applying an integral image. This can be summarized as a series of processes to find out the structure of the distribution of the CS magnitudes contained in the image when the calculated CS is regarded as one image.

In this experiment, we expected the CNN to be able to learn features without manually extracting feature extraction from CS images. Convolution calculates the correlation between an image and a filter, which acts to detect where the image resembles the pattern of the filter. The convolutional layer used in CNN is a single layer neural network that performs this convolution operation. In the convolution operation, a local coupling of the input layer and the output layer occurs, and the weight of the coupling is a coefficient of the filter. Since this weight is the filter itself to know the pattern of the presented image, it can be summarized as the process of optimizing this filter through CNN for the classification task. Therefore, superior results of CNN mean that CNN is better able to recognize the CS image pattern difference of the snoring sound than the hand-crafted method.

We also used this newly generated sleep breathing sound classifier as a feature extractor in the OSA severity classification task. Based on the same dataset used in Chapter 4, we used a temporal analysis method proposed in the same chapter to extract a short ROI of about 10 minutes. A new feature representation was extracted from the partial respiration sound through the SRBD-related snoring sound classifier described above, and a classification model was created using the extracted feature representation. Despite using a relatively short sleep breathing interval, it showed better classification performance than previous studies. Especially, the relatively low specificity result (75%), which was a problem of the existing study results in the OSA detection experiment, is greatly improved in the newly proposed framework. Specificity is an indicator of whether normal patients can be distinguished as a normal patient. Its low value can be a major reason for users to underestimate the reliability of devices and associated algorithms in personal healthcare services. Because the OSA detection experiment was performed based on the AHI value of 5 as in the previous experiment, it is classified as the OSA patient from the mild OSA patient. In this case, the average sensitivity that can show the performance distinguishing the patients having the OSA symptom correctly is about 91 % (in the case using an output of the CNN's dense layer). This shows that the sensitivity is slightly lower when compared to previous experiments using full breathing sounds. Because there is not a large number of data involved in the experiment, a few classification errors can lead to significant differences in results. However, as a result, the improvement in specificity will compensate for the decline of sensitivity.

Of course, as in previous experiments, it is not an experiment using large-scale patient data, so the generalized error of the developed classifier may still exist. However, it is evident from objective experiments that the learned features perform better than the hand-crafted features when using the CS derived from sleep breathing sounds of the same data set.

In addition, we can confirm the performance improvement by compounding existing hand-crafted features on learned features. Therefore, it is expected that the better classification performance will be achieved if the feature learning method and the hand-crafted method are appropriately mixed or combined in the related studies. In this test, however, this is limited to hand-crafted features extracted from the previous full breathing sounds. The transition probability, a hand-crafted feature calculated from the 10-minutes ROI extracted in this experiment, did not affect the performance improvement. This means that the simple amplitude value or time interval information of the signal, which is not able to represent the characteristics of the sound chunk itself, cannot properly reflect its characteristics given short-length data.

In this study, we can confirm that the feature learning method can be effective in generating an effective feature set if some specific data form which is transformed from raw data shows a complicated data distribution, or if the interpretation is relatively complex. Besides, this study extracted OSA severity of individual patients by extracting short 10 minutes sleep breathing sounds from average four hours total sleep breathing sounds. This type of analysis framework can save a significant amount of computing power which is required for large data analysis when it applied to personal healthcare services.

5.5 Summary

In this study, SRBD-related snoring classification and OSA severity classification tasks are performed using the latest deep learning technology. Based on CS computed from snoring samples, it is at the core of this experiment that snore sound samples can be distinguished by symptoms through the feature learning. As far as we know, this is the first experiment to use deep learning in the study of sleep breathing sounds including snoring. This experiment can also be compared with the study of Chapter 3 of this thesis. As a result, CNN-based features showed better performance in classification experiments using more data sets than COV-II (covariance integral image of CS)-based hand-crafted features, which showed the best performance in Chapter 3. It is considered that the feature learned through CNN better describes the structure of the image which indicates the distribution of CS magnitude values matching the spectral frequency value. However, still, without a standardized experimental framework, it is hard to collect data and to explain the classification performance completely objectively. Therefore, if standardized data collection and experimental framework are developed in the future, it is expected that a significant amount of data and various deep learning structures, including our method, will generate a reliable classifier for the SRBD-related snoring sound that can minimize generalization error. Also, we performed the previous OSA severity classification task using this classifier. In particular, we assume that the SRBD-related snoring classifier has a sufficiently good performance. Thus, only the 10-minutes interval with the greatest energy change in the total breathing sounds was selected experimentally from existing sound dataset for OSA classification. The results of the experiment showed better results, especially the improvement of the specificity, than the previous experiments. The feature used in this experiment is the output derived from specific layers of the CNN structure for the SRBD-related snoring classifier, and the severity class was identified using a general classifier such as SVM. The fact that the experiment in the short time interval shows better performance than the previous experiment means that SRBD-related snoring classifier using CNN extracts better features from snoring sound samples. However, this experiment is expected to have generalization error because OSA classification is performed using a large number of CNN's parameters as a feature based on a small number of patient data. Therefore, as mentioned above, we expect to be able to develop a more robust OSA severity classifier if we can utilize large patient datasets in the analysis and classification framework developed in this experiment, and we also expect that our proposed framework can be used for various personal healthcare services using sleep breathing sounds.

Chapter 6. Conclusions and Future Work

6.1 Conclusions

The purpose of this study was to classify the SRBD-related snoring sounds and severity of obstructive sleep apnea of patients using only breathing sounds during sleep. We extracted various features from sleep breathing sounds recorded during PSG using different analysis methods, and performed SRBD-related snoring classification, and OSA severity classification of OSA suspected patients. The analytical framework derived from this study is expected to be used in the future for sleep breathing sound research, screening for PSG testing, and home healthcare services. In particular, since sleep breathing sounds can be recorded without any other additional sensors, it will be possible to perform necessary tasks with various mobile or wearable devices in the future without high-cost investment for the related functions. In terms of the results provided, this study can assess the patient's OSA severity in more detail than other studies. Therefore, more detailed information can be provided to the patient or physician when the actual PSG screening test is carried out so that it can be utilized as a differentiated service.

Many studies using breathing sounds, including snoring, have established a professional sound recording environment and have placed a microphone in a position very close to the patient or attached a microphone to his/her body. In this thesis, we used the sound extracted from the recorded video using the microphone and camcorder installed relatively far from the patient (approximate 2 meter) to monitor the test environment including the patient in the PSG test. Although not completely consistent with the user's actual environment, by using a sound similar to a typical recording environment as input, we can focus more on the extraction and development of features available in the normal environment. Also, it is possible to exclude the serious limitations that may be raised in actual use, which may be caused by the location of the microphone or the type of the microphone.

However, this experiment cannot be applied directly to the latest smart handheld devices even if a professional recording environment and a special microphone are not used. The microphone used in this study is still installed on the PSG room ceiling of the hospital. Also, experiments are not carried out using the sleep breathing sounds recorded with the microphones built in the user's smart device at practically various locations. However, as described in Chapter 4 of the system architecture, this study is based on experimentation with a sound that is worse than that recorded in a user's ordinary recording environment. This could be advantageous when applied to a general user environment or device rather than a conventional experiment in which various conditions must be accompanied.

In a view of the patient data analysis system for expansion into various health care environments, the significance of this thesis is that it has confirmed the feasibility of a system capable of analyzing respiratory sounds associated with sleeping breathing disorder and OSA severity classification using only sleep breathing sounds with the devices and recording level and conditions that can be considered by the patient. To expand and evolve the analysis framework of this experiment to a substantial home healthcare and mobile healthcare environment, some additional tasks should be considered, and these are the limitations of this study.

First, various sounds other than sleep breathing sounds should be able to be filtered in the user's general environment. Considering services in general smart devices, it could be possible to record selective sleep breathing sounds through motion detection using a built-in sensor (for example, a 3-axis accelerometer) or prediction of a simple sleep phase. However, depending on the recording environment or the device used, it may be necessary to filter out noise that may degrade the quality of the sleep breathing sounds. Moreover, the environment in which more than two patients simultaneously develop snoring sound can also be considered, which is expected to require a high level of filtering because of the need for individual discrimination of sleep breathing sounds.

Second, a standardized experimental protocol and framework are required. These requires considerable time and effort. However, the gold standard for processing individual data is essential for analyzing respiratory sounds associated with various sleep disorders. In this thesis, the annotation related to the symptom of sleep disorder diagnosed by the physician in the sleep analysis system used in the PSG test is taken as the gold standard, and the respiration sound in the matching time interval was extracted as data representative of the specific symptom. However, since this annotation does not match with the experimental protocol for analyzing sleep breath sounds, it is hard to extract related breathing sounds, and various assumptions are needed (For example, an OSA-related annotation is matched to silence interval from sleep breathing noise, hypopnea is different from diagnosis interval and snoring period). Therefore, a standardized experimental framework is necessary if robust and accurate experiments and scalability are considered.

Third, several types of microphones and different recording positions should be considered. In an actual user's healthcare service environment, the type and recording location of the microphone may affect the performance of the analysis and classification algorithms. A variety of experiments are needed to find optimal conditions, but they should not undermine the ultimate goals such

Figure 6.1 Main contributions of this research

as patient custom or mobile healthcare due to unrealistic configurations.

Finally, high-quality sleep breathing sounds are still required for the use of sleep breath sounds in a variety of studies or applications. Although, the studies to detect the occurrence and severity of certain symptoms using breathing sound are important, In the future, it will be a major research goal to predict the possible diseases or conditions in the future by using specific information inherent in breathing sounds. Therefore, to be able to do such research, we should be able to discover features that can predict the correlation with various diseases by using high-quality sleep breathing sounds.

Nevertheless, this research presented a new analytical method for analyzing snoring and breathing sounds during sleep. Based on the fact that snoring is produced not from the vocal cords but by various anatomical organs including the tongue, larynx, and soft palate, we have proposed a classification method based on cyclostationary analysis rather than general acoustical analysis methods. Although there is no formal database or evaluation criteria for the proposed method, it is difficult to compare the performance of other similar studies directly, but it would be appropriate to use it for the screening test of general obstructive sleep apnea.

The contribution of this study can be summarized as the following items, and the related studies were conducted in this regard.

① By using only sleep breathing sounds, it is possible to eliminate inconveniences caused by wearing the biomedical sensors, and effectively respond to patient custom or mobile healthcare services.

② By extending OSA detection studies, which were performed mainly by previous studies, to multiclass OSA severity classification research based on AHI value, more detailed information can be provided to users in actual healthcare service.

③ A novel analysis method for sleep breathing sounds using cyclic spectrum, energy signal transition probability and deep learning technique was applied through this study.

In the SRBD-related snoring analysis, we attempted to classify simple snoring, OSA-related snoring, and hypopnea-related snoring. The AHI, which indicates OSA severity, is calculated based on the total number of apnea and hypopnea occurring during the entire sleep period. In the result, we could not classify three types of snoring very clearly but could suggest the possibility that the snoring associated with apnea symptoms could be classified using only breathing sounds. Therefore, we thought that OSA severity could be judged only by analysis of snoring events if even snoring associated with the hypopnea, which is evaluated based on changes in oxygen saturation and a time interval of respiration stop, can be classified directly using sounds. This experiment was a core study of features extraction from breathing sounds, and used various hand-crafted feature extraction methods and deep learning technique-based feature learning for related classification experiments. We conducted several studies in parallel and extended the experimental scope based on the study's goal. First of all, we proposed a model in which a given snoring sound is classified into simple snoring, hypopnea related snoring, and apnea related snoring based on the recorded sleep breathing sounds from PSG. Based on the various SRBDs checked by the medical staff at PSG, snoring sounds were extracted, and cyclostationary analysis was applied to each sound. From the cyclic spectrum (CS) function obtained as a result of this analysis, we derive the final feature set through dimension reduction, statistical analysis, and feature selection. Through the proposed method, we could confirm by experiment that each SRBD-related snoring sound has different characteristics. This demonstrates the possibility of estimating the relevant SRBD through the snoring sound analysis. In the related experiments described in Chapter 3, three SRBD-related snoring sounds classifiers showed performance of 73.8 ~ 78.07 %.

Secondly, we performed a cyclostationary and temporal analysis of individual sleep breathing sounds, without detecting specific events in each section, and obtained feature representations reflecting the key property of sleep breathing sounds of a patient. Temporal analysis extracted the feature representing energy transition probability in the sleep breathing sound, which represented the probability that a particular energy level of a breathing sound changes to a different energy level in the time domain. The OSA suspected interval was set using the information on the silence interval, and the transition probability for this interval was also calculated at the same time. Cyclostationary analysis was performed by calculating the CS by each segment of the sleep breathing sounds, followed by dimension reduction, statistical processing, and feature selection. Through the feature set derived from these two methods, the patient's sleep breathing sounds were classified into four AHI severity categories with 79.52 % accuracy. Due to the absence of a standard test framework, it is difficult to directly compare performance with other studies, but the results of this experiment are relatively high classification performance in the same classification task.

Finally, we have advanced the algorithm by applying the deep learning technology to the existing analysis method. Based on CNN, we attempted feature learning in CS and classify the sleep breathing sound event of the patient using the derived parameters and confirmed the performance. The SRBD-related snoring classifier obtained through this experiment was applied to the existing OSA severity classification task as a feature extractor. Especially, in OSA severity discrimination experiment using deep learning, partial sleep breathing sounds were extracted by recognizing the part with relatively large variance in total breathing sounds. We extracted features in this interval and estimated the OSA severity of the patient. In this experiment, we could confirm the best classification performance of 85 % in the OSA severity classification based on deep learning technology using CNN's fully connected layer output. This result shows higher performance than the classification experiment using the previous hand-crafted based feature extraction method.

In this thesis, the sound derived from human body was applied to task related to sleep disorder using machine learning technique. To apply machine learning to a specific task, a significant amount of data is required. However, data collection for healthcare related tasks is often more challenging than other data due to privacy and user safety issues. It is therefore essential to cooperate with institutions or hospitals that enable continuous data availability. Nevertheless, if the number of participants is not large, a classification experiment based on the detection of a specific sound event justly may be more efficient. For this, it is important to obtain high-quality sound from the patient and to study the common characteristics of the sound associated with the specific symptom.

Besides, in the case of healthcare-related experiments using sounds, the creation of a gold standard is of utmost importance, although it has been mentioned several times in the preceding. Clearly, pre-classified sound samples with specific symptoms are required. However, since the diagnostic results are not often generated based on the sound, most of them will use the acquired sound file in synchronization with the actual diagnostic parameters in time. At this time, it is preferable from the viewpoint of expansion of experiments, making an experimental protocol capable of extracting the sample on a clear criterion.

In this thesis, we have found that sleep breathing sound can be used to identify the conditions that have a fatal effect on a patient's sleep health. If we can analyze the sleep breathing sound and various environmental sounds at the same time, it will be possible to provide various healthcare services such as sleep health monitoring, and emergency status detection of patients in a bedroom environment without any additional sensors. Also, if the actual target device and the recording environment to record are confirmed, it will be possible to analyze the respiration sound in more detail and expand the related research to various aspects according to specific purposes.

6.2 Future work

In the future study, we will seek to improve generalization errors and improve accuracy by improving the analysis framework based on data collected from more patients through large-scale research. There is also a need to improve the reliability and robustness of the analysis framework by performing classification studies on various sounds, including snoring sounds as well as specific noises that occur during sleep. Of course, future extension studies need a clear specification of system setup related to sound acquisition. Therefore, large-scale data collection and analysis should be performed with the type of microphone, target device, and recording location clearly set according to the application purpose.

Based on the above extended studies, we would like to apply this analysis framework to mobile healthcare services related to sleep health. However, for mobile healthcare services, there are a number of additional considerations that need to be considered along with the performance of the analysis framework itself. The power consumption of a target device and execution time for analysis and providing results are very important constraints to be considered in the mobile environments. Therefore, it is necessary to develop and operate an efficient analysis algorithm considering the computation ability of the target device. In this respect, research on the interaction between machines and humans in the field of sleep related healthcare is also an important topic of interest. Finding methods to effectively provide collected data or analytical results to users or physicians is critical to the efficiency and user-friendly application of the overall system.

Identifying the relationship between the patient's voice and OSA severity is also an interesting topic for future research. As mentioned earlier, speech and snoring sounds have different development mechanisms. However, deformation of the UA structure due to aging and obesity can affect voice, and this change may be characterized by OSA severity and recorded for a specific trend analysis. For this study, it is necessary to observe very specific changes associated with symptoms from the sound. Therefore, high-quality recorded voices are required and the development of sentences or pronunciation that can well reflect the sound characteristics associated with pathological symptoms may be needed.

The standard experimental protocol and the sleep breathing sound database are indispensable factors in future studies. These factors enable objective evaluation of the algorithm under development and benchmarking for the performance improvement. In the future, we would also like to conduct research standardization work for various sleep breathing sounds analysis studies in conjunction with related researchers and hospitals. To this end, synchronization of the PSG result data with sleep breathing sounds, construction of related storage servers, and development of data extraction tools based on these components could be significant starting points for the standard experiment protocol.

Bibliography

1. Guadagni V, Burles F, Ferrara M, Iaria G: **The effects of sleep deprivation on emotional empathy.** *J Sleep Res* 2014, **23**(6):657-663.
2. van der Helm E, Gujar N, Walker MP: **Sleep deprivation impairs the accurate recognition of human emotions.** *Sleep* 2010, **33**(3):335-342.
3. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA: **Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies.** *Eur Heart J* 2011, **32**(12):1484-1492.
4. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA: **Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis.** *Diabetes Care* 2010, **33**(2):414-420.
5. Gottlieb DJ, Punjabi NM, Newman AB, Resnick HE, Redline S, Baldwin CM, Nieto FJ: **Association of sleep time with diabetes mellitus and impaired glucose tolerance.** *Arch Intern Med* 2005, **165**(8):863-867.
6. Rosekind MR, Gregory KB, Mallis MM, Brandt SL, Seal B, Lerner D: **The cost of poor sleep: workplace productivity loss and associated costs.** *J Occup Environ Med* 2010, **52**(1):91-98.
7. Sagberg F: **Road accidents caused by drivers falling asleep.** *Accid Anal Prev* 1999, **31**(6):639-649.
8. Stutts JC, Wilkins JW, Scott Osberg J, Vaughn BV: **Driver risk factors for sleep-related crashes.** *Accid Anal Prev* 2003, **35**(3):321-331.
9. Thorpy MJ: **Classification of sleep disorders.** *Neurotherapeutics* 2012, **9**(4):687-701.
10. Earley CJ: **Clinical practice. Restless legs syndrome.** *N Engl J Med* 2003, **348**(21):2103-2109.
11. Aldrich MS: **Narcolepsy.** *Neurology* 1992, **42**(7 Suppl 6):34-43.
12. Fry JM, DiPhillipo MA, Curran K, Goldberg R, Baran AS: **Full polysomnography in the home.** *Sleep* 1998, **21**(6):635-642.
13. Zou D, Grote L, Peker Y, Lindblad U, Hedner J: **Validation a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography.** *Sleep* 2006, **29**(3):367-374.
14. Epstein LJ, Dorlac GR: **Cost-effectiveness analysis of nocturnal oximetry as a method of screening for sleep apnea-hypopnea syndrome.** *Chest* 1998,

113(1):97-103.

15. Mendels J, Hawkins DR: **Sleep laboratory adaptation in normal subjects and depressed patients ("first night effect").** *Electroencephalogr Clin Neurophysiol* 1967, **22**(6):556-558.
16. Williams A, Santiago S, Stein M: **Screening for sleep apnea?** *Chest* 1989, **96**(3):451-453.
17. Fiz JA, Abad J, Jane R, Riera M, Mananas MA, Caminal P, Rodenstein D, Morera J: **Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea.** *Eur Respir J* 1996, **9**(11):2365-2370.
18. Ng AK, Koh TS, Baey E, Lee TH, Abeyratne UR, Puvanendran K: **Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?** *Sleep Med* 2008, **9**(8):894-898.
19. Perez-Padilla JR, Slawinski E, Difrancesco LM, Feige RR, Remmers JE, Whitelaw WA: **Characteristics of the snoring noise in patients with and without occlusive sleep apnea.** *Am Rev Respir Dis* 1993, **147**(3):635-644.
20. Jung R, Kuhlo W: **Neurophysiological Studies of Abnormal Night Sleep and the Pickwickian Syndrome.** *Prog Brain Res* 1965, **18**:140-159.
21. He J, Kryger MH, Zorick FJ, Conway W, Roth T: **Mortality and apnea index in obstructive sleep apnea. Experience in 385 male patients.** *Chest* 1988, **94**(1):9-14.
22. Wang H, Parker JD, Newton GE, Floras JS, Mak S, Chiu KL, Ruttanaumpawan P, Tomlinson G, Bradley TD: **Influence of obstructive sleep apnea on mortality in patients with heart failure.** *J Am Coll Cardiol* 2007, **49**(15):1625-1631.
23. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ: **The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.** *Psychiatry Res* 1989, **28**(2):193-213.
24. Weatherwax KJ, Lin X, Marzec ML, Malow BA: **Obstructive sleep apnea in epilepsy patients: the Sleep Apnea scale of the Sleep Disorders Questionnaire (SA-SDQ) is a useful screening instrument for obstructive sleep apnea in a disease-specific population.** *Sleep Med* 2003, **4**(6):517-521.
25. Johns MW: **A new method for measuring daytime sleepiness: the Epworth sleepiness scale.** *Sleep* 1991, **14**(6):540-545.
26. Williams AJ YG, Santiago S, Stein M: **Screening for Sleep Apnea Using**

Pulse Oximetry and A Clinical Score. *Chest* 1991, **100**:631-635.

27. Erman MK, Stewart D, Einhorn D, Gordon N, Casal E: **Validation of the ApneaLink for the screening of sleep apnea: a novel and simple single-channel recording device.** *J Clin Sleep Med* 2007, **3**(4):387-392.
28. Mendez MO, Bianchi AM, Matteucci M, Cerutti S, Penzel T: **Sleep apnea screening by autoregressive models from a single ECG lead.** *IEEE Trans Biomed Eng* 2009, **56**(12):2838-2850.
29. Roche F, Gaspoz JM, Court-Fortune I, Minini P, Pichot V, Duverney D, Costes F, Lacour JR, Barthelemy JC: **Screening of obstructive sleep apnea syndrome by heart rate variability analysis.** *Circulation* 1999, **100**(13):1411-1415.
30. Morillo D, Rojas Ojeda JL, Crespo Foix LF, Jimenez A: **An accelerometer-based device for sleep apnea screening.** *IEEE Trans Inf Technol Biomed* 2010, **14**(2):491-499.
31. Sola-Soler J, Jane, R., Fiz, J. A., & Morera, J.: **Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea.** *Engineering in Medicine and Biology Society, 2003 Proceedings of the 25th Annual International Conference of the IEEE* 2003, **3**:2527-2530.
32. Strollo PJ, Jr., Rogers RM: **Obstructive sleep apnea.** *N Engl J Med* 1996, **334**(2):99-104.
33. Strohl KP, Redline S: **Recognition of obstructive sleep apnea.** *Am J Respir Crit Care Med* 1996, **154**(2 Pt 1):279-289.
34. Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM: **The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index.** *J Clin Sleep Med* 2011, **7**(5):459-465B.
35. Madisetti VK: *The digital signal processing handbook.* CRC Press; 2009.
36. Lee LA, Lo YL, Yu JF, Lee GS, Ni YL, Chen NH, Fang TJ, Huang CG, Cheng WN, Li HY: **Snoring Sounds Predict Obstruction Sites and Surgical Response in Patients with Obstructive Sleep Apnea Hypopnea Syndrome.** *Sci Rep* 2016, **6**:30629.
37. Hill PD, Lee BW, Osborne JE, Osman EZ: **Palatal snoring identified by acoustic crest factor analysis.** *Physiol Meas* 1999, **20**(2):167-174.
38. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521**(7553):436-

444.
39. Krizhevsky A, Sutskever I, Hinton GE: **Imagenet classification with deep convolutional neural networks.** In *Advances in neural information processing systems*. 2012:1097-1105.
40. Platt J: **Sequential minimal optimization: A fast algorithm for training support vector machines.** 1998.
41. Crocker MJ: *Handbook of acoustics.* John Wiley & Sons; 1998.
42. Pevernagie D, Aarts RM, De Meyer M: **The acoustics of snoring.** *Sleep Med Rev* 2010, **14**(2):131-144.
43. Osborne JE, Osman EZ, Hill PD, Lee BV, Sparkes C: **A new acoustic method of differentiating palatal from non-palatal snoring.** *Clin Otolaryngol Allied Sci* 1999, **24**(2):130-133.
44. Kim J, Kim T, Lee D, Kim J-W, Lee K: **Exploiting temporal and nonstationary features in breathing sound analysis for multiple obstructive sleep apnea severity classification.** *Biomedical engineering online* 2017, **16**(1):6.
45. Abeyratne UR, Wakwella AS, Hukins C: **Pitch jump probability measures for the analysis of snoring sounds in apnea.** *Physiol Meas* 2005, **26**(5):779-798.
46. Matsiki D, Deligianni X, Vlachogianni-Daskalopoulou E, Hadjileontiadis LJ: **Wavelet-based analysis of nocturnal snoring in apneic patients undergoing polysomnography.** *Conf Proc IEEE Eng Med Biol Soc* 2007, **2007**:1912-1915.
47. Logan B: **Mel Frequency Cepstral Coefficients for Music Modeling.** In *ISMIR*. 2000.
48. Marupaka PT, Singh RK: **Comparison of classification results obtained by using cyclostationary features, MFCC, proposed algorithm and development of an environmental sound classification system.** In *Advances in Electronics, Computers and Communications (ICAECC), 2014 International Conference on*. IEEE; 2014:1-6.
49. Quinn S, Daly N, Ellis P: **Observation of the mechanism of snoring using sleep nasendoscopy.** *Clinical Otolaryngology* 1995, **20**(4):360-364.
50. Radhi N, Aziz K, Abbas S, AL-Raweshidy H: **Cyclostationary detection in spectrum pooling system of undefined secondary users.** In *The Seventh*

International Conference on Wireless and Mobile Communications, Luxembourg, Belgium. 2011:266-270.

51. Khalaf AF, Owis MI, Yassine IA: **A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines.** *Expert Syst Appl* 2015, **42**(21):8361-8368.
52. Quinn SJ, Huang L, Ellis PD, Williams JEF: **The differentiation of snoring mechanisms using sound analysis.** *Clinical Otolaryngology* 1996, **21**(2):119-123.
53. Agrawal S, Stone P, McGuinness K, Morris J, Camilleri AE: **Sound frequency analysis and the site of snoring in natural and induced sleep.** *Clin Otolaryngol Allied Sci* 2002, **27**(3):162-166.
54. Jones T, Ho M, Earis J, Swift A, Charters P: **Acoustic parameters of snoring sound to compare natural snores with snores during 'steady-state'propofol sedation.** *Clinical Otolaryngology* 2006, **31**(1):46-52.
55. Schafer J, Pirsig W: **Digital signal analysis of snoring sounds in children.** *Int J Pediatr Otorhinolaryngol* 1990, **20**(3):193-202.
56. Saunders N, Tassone P, Wood G, Norris A, Harries M, Kotecha B: **Is acoustic analysis of snoring an alternative to sleep nasendoscopy?** *Clinical Otolaryngology* 2004, **29**(3):242-246.
57. Herzog M, Schmidt A, Bremert T, Herzog B, Hosemann W, Kaftan H: **Analysed snoring sounds correlate to obstructive sleep disordered breathing.** *Eur Arch Otorhinolaryngol* 2008, **265**(1):105-113.
58. Ng AK, Wong KY, Tan CH, Koh TS: **Bispectral analysis of snore signals for obstructive sleep apnea detection.** *Conf Proc IEEE Eng Med Biol Soc* 2007, **2007**:6196-6199.
59. Bertino G, Matti E, Migliazzi S, Pagella F, Tinelli C, Benazzo M: **Acoustic changes in voice after surgery for snoring: preliminary results.** *Acta Otorhinolaryngol Ital* 2006, **26**(2):110-114.
60. Lee H, Pham P, Largman Y, Ng AY: **Unsupervised feature learning for audio classification using convolutional deep belief networks.** In *Advances in neural information processing systems*. 2009:1096-1104.
61. Kim J, Kim T, Lee K, Kim J-W: **CYCLOSTATIONARITY-BASED ANALYSIS OF NOCTURNAL RESPIRATION SOUNDS.**
62. Nakano H, Hayashi M, Ohshima E, Nishikata N, Shinohara T: **Validation of**

a new system of tracheal sound analysis for the diagnosis of sleep apnea-hypopnea syndrome. *Sleep* 2004, **27**(5):951-957.

63. Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, Stradling J, Clifford GD: **SleepAp: An Automated Obstructive Sleep Apnoea Screening Application for Smartphones.** *Ieee J Biomed Health* 2015, **19**(1):325-331.

64. Azarbarzin A, Moussavi Z: **Snoring sounds variability as a signature of obstructive sleep apnea.** *Med Eng Phys* 2013, **35**(4):479-485.

65. Sola-Soler J, Jane R, Fiz J, Morera J: **Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea.** In *Engineering in Medicine and Biology Society, 2003 Proceedings of the 25th Annual International Conference of the IEEE*. IEEE; 2003:2527-2530.

66. McNicholas W, Bonsignore M, B26 MCoECA: **Sleep apnoea as an independent risk factor for cardiovascular disease: current evidence, basic mechanisms and research priorities.** *European Respiratory Journal* 2007, **29**(1):156-178.

67. Ben-Israel N, Tarasiuk A, Zigel Y: **Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults.** *Sleep* 2012, **35**(9):1299-1305C.

68. **FFmpeg. 2012.** [https://www.ffmpeg.org]

69. Boll S: **Suppression of acoustic noise in speech using spectral subtraction.** *IEEE Transactions on acoustics, speech, and signal processing* 1979, **27**(2):113-120.

70. Priestley MB: **Evolutionary spectra and non-stationary processes.** *Journal of the Royal Statistical Society Series B (Methodological)* 1965:204-237.

71. Gardner WA, Napolitano A, Paura L: **Cyclostationarity: Half a century of research.** *Signal processing* 2006, **86**(4):639-697.

72. Lima AF: **Analysis of low probability of intercept (LPI) radar signals using cyclostationary processing.** Monterey California. Naval Postgraduate School; 2002.

73. Tuzel O, Porikli F, Meer P: **Pedestrian detection via classification on riemannian manifolds.** *IEEE transactions on pattern analysis and machine intelligence* 2008, **30**(10):1713-1727.

74. Tuzel O, Porikli F, Meer P: **Region covariance: A fast descriptor for**

detection and classification. *Computer Vision–ECCV 2006* 2006:589-600.

75. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH: **The WEKA data mining software: an update.** *ACM SIGKDD explorations newsletter* 2009, **11**(1):10-18.

76. Zhu W, Zeng N, Wang N: **Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations.** *NESUG proceedings: health care and life sciences, Baltimore, Maryland* 2010:1-9.

77. Muller MP, Tomlinson G, Marrie TJ, Tang P, McGeer A, Low DE, Detsky AS, Gold WL: **Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia?** *Clinical infectious diseases* 2005, **40**(8):1079-1086.

78. Flach P: *Machine learning: the art and science of algorithms that make sense of data.* Cambridge University Press; 2012.

79. Tregear S, Reston J, Schoelles K, Phillips B: **Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis.** *J Clin Sleep Med* 2009, **5**(6):573-581.

80. Gilani S, Quan SF, Pynnonen MA, Shin JJ: **Obstructive sleep apnea and gastroesophageal reflux: a multivariate population-level analysis.** *Otolaryngology--Head and Neck Surgery* 2016, **154**(2):390-395.

81. Kozlowska B, Wydrych K, Wydrych P, Minarowski L, Chwiesko-Minarowska S, Kuryliszyn-Moskal A, Chyczewska E: **Libido disorders in patients with obstructive sleep apnea.** *European Respiratory Journal* 2014, **44**(Suppl 58):P4923.

82. Guilleminault C, Connolly SJ, Winkle RA: **Cardiac arrhythmia and conduction disturbances during sleep in 400 patients with sleep apnea syndrome.** *The American journal of cardiology* 1983, **52**(5):490-494.

83. Young T, Peppard PE, Gottlieb DJ: **Epidemiology of obstructive sleep apnea: a population health perspective.** *Am J Respir Crit Care Med* 2002, **165**(9):1217-1239.

84. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM: **Increased prevalence of sleep-disordered breathing in adults.** *Am J Epidemiol* 2013, **177**(9):1006-1014.

85. Young T, Evans L, Finn L, Palta M: **Estimation of the clinically diagnosed**

proportion of sleep apnea syndrome in middle-aged men and women. *Sleep* 1997, **20**(9):705-706.

86. Medicine AAoS: **International Classification of Sleep Disorders, Revised: Diagnostic and Coding Manual. Chicago, IL: American Academy of Sleep Medicine; 2001.** In; 2013.

87. Punjabi NM: **The epidemiology of adult obstructive sleep apnea.** *Proc Am Thorac Soc* 2008, **5**(2):136-143.

88. Stevens DR: *Sleep medicine secrets.* Elsevier Health Sciences; 2004.

89. Jones T, Swift A, Calverley P, Ho M, Earis J: **Acoustic analysis of snoring before and after palatal surgery.** *European Respiratory Journal* 2005, **25**(6):1044-1049.

90. Randall RB, Antoni J, Chobsaard S: **The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals.** *Mechanical systems and signal processing* 2001, **15**(5):945-962.

91. Otsu N: **A threshold selection method from gray-level histograms.** *IEEE Transactions on systems, man, and cybernetics* 1979, **9**(1):62-66.

92. Lee DD, Seung HS: **Algorithms for non-negative matrix factorization.** In *Advances in neural information processing systems*. 2001:556-562.

93. Gutlein M, Frank E, Hall M, Karwath A: **Large-scale attribute selection using wrappers.** In *Computational Intelligence and Data Mining, 2009 CIDM'09 IEEE Symposium on*. IEEE; 2009:332-339.

94. Pudil P, Novovičová J, Kittler J: **Floating search methods in feature selection.** *Pattern recognition letters* 1994, **15**(11):1119-1125.

95. Cristianini N, Scholkopf B: **Support vector machines and kernel methods: the new generation of learning machines.** *Ai Magazine* 2002, **23**(3):31.

96. Maglogiannis I, Loukis E, Zafiropoulos E, Stasis A: **Support vectors machine-based identification of heart valve diseases using heart sounds.** *Computer methods and programs in biomedicine* 2009, **95**(1):47-61.

97. Noble WS: **What is a support vector machine?** *Nature biotechnology* 2006, **24**(12):1565-1567.

98. Hinton GE, Osindero S, Teh YW: **A fast learning algorithm for deep belief nets.** *Neural Comput* 2006, **18**(7):1527-1554.

99. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ:

Performance-optimized hierarchical models predict neural responses in higher visual cortex. *P Natl Acad Sci USA* 2014, **111**(23):8619-8624.

100. Chollet Fc, ois, others: **Keras.** In: \url{https://github.com/fchollet/keras}; 2015.

101. Group H: **Hierarchical data format, version 5.** 2014.

102. Kohavi R, John GH: **Wrappers for feature subset selection.** *Artif Intell* 1997, **97**(1-2):273-324.

초 록

폐쇄성 수면 무호흡증 (Obstructive Sleep Apnea, OSA) 은 대표적인 수면 질환으로 유병률이 높으며 고혈압 및 뇌졸중의 위험 요소로서 사망률을 증가시킨다. 이 증상은 수면 중에 발생하므로 환자가 자각하기 어렵고, 수면다원검사라는 표준검사가 있으나 복잡한 진단 방식과 고비용에 대한 부담으로 인하여 실제 진단율은 낮은 편이다. 따라서 해당 증상 의심 환자에 대한 수면다원검사의 필요 여부를 판단할 수 있는 효과적이고 합리적인 선별 검사에 대한 요구가 증가하고 있다. 본 논문에서는 수면 중 호흡음만을 이용하여 환자의 코골이음 및 OSA 중증도를 분류하는 세가지 연구를 수행하였다. 먼저 주기 정상성 분석에 기초한 특징점을 이용하여 수면 호흡장애 관련 코골이음의 분류 가능성을 확인하였다. 그리고 장시간의 수면 호흡음으로부터 시간 및 주기 정상성 분석에 기반한 특징점들을 추출하고 환자들의 OSA 중증도를 분류하였다. 최종적으로는 상기 분류 태스크들의 효율성과 성능 향상을 위하여 부분 수면 호흡음 추출 및 컨볼루션 신경망을 이용한 특징 학습 과정을 실험에 적용하였다. 컨볼루션 신경망을 이용한 수면 호흡음 분석 방법은 다중 클래스 코골이음 및 OSA 중증도 분류 태스크에서 80 % 이상의 분류 정확도, 평균 AUC (Area Under Curve) > 0.8 이상의 우수한 성능을 보여주었다. 제안된 방법은 향후 환자 맞춤형 헬스케어 서비스에서 수면다원검사의 효율성 및 정확성 향상을 위한 선별 검사의 핵심 분석 도구로 활용될 수 있을 것으로 기대한다.

주요어: 코골이 소리 분석, 폐쇄성 수면 무호흡, 선별 검사, 주기 정상성 분석, 특징 학습, 컨벌루션 신경망, 중증도 분류

학 번: 2013-31259