Obstructive Sleep Apnea Diagnosis With Apnea Event Detection in Snoring Sound Using a Conditional Random Field

He Lian (A0068205J)

B.Sc. Of Computer Science

Peking University

2010

A Thesis Submitted

For The Degree of Master of Science

Department of Computer Science

School of Computing

National University of Singapore

2012

Abstract

Obstructive Sleep Apnea (OSA) has become increasingly prevalent throughout the world in recent decades, but its proper diagnosis is severely constrained by the limited accessibility of polysomnography (PSG) facilities. To resolve this problem, researchers investigated the potential of OSA diagnosis by using snore-related sounds. However, most existing approaches to OSA diagnosis analyze snore episodes or silence episodes individually. In this thesis, we propose a method to identify apnea events by incorporating ISPJ and F1 lables and learning the relation among these sequential acoustic signal components using a conditional random field. Compared with three existing methods, the proposed method exhibits the best performance by achieving a sensitivity of 92.31% and a specificity of 80% under the threshold of apnea index set to 5. Moreover, the number of apnea events detected by our approach effectively approximates the actual one reported by PSG, which makes the proposed method a potential alternative for manual annotation. Based on the proposed method, a prototype named Mobile Obstructive Sleep Apnea Diagnosis is implemented on a mobile device. Validation results demonstrate the prototype's effectiveness and efficiency. The efficacy and portability of our system illustrate its promising potential for OSA screening in a home environment.

Acknowledgment

I would like to express my sincere gratitude to Dr. Wang Ye for his guidance and encouragement, to Dr. Sim Khe Chai for his generous support, and to Dr. Khoo See Meng for his cooperation in the data collection. I am also grateful to Lee Yue Ting and Liu Liu for their assistance. Lastly, I would like to show my appreciation to Fang Haotian for proofreading my thesis, as well as to everyone who has helped me along the way. Thank you.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Obstructive Sleep Apnea (OSA) is the most common sleep-related breathing disorder. It is characterized by the total or partial obstruction of the upper airway during sleep, accompanied by repetitive cessation of respiratory airflow and frequent premature arousals. Untreated OSA reduces the quality of sleep and increases the risk of heart disease, cognitive impairment, high blood pressure and stroke. The loss of restorative sleep causes sleepiness during the day and contributes to the rising number of motor accidents [18, 38].

OSA has become increasingly prevalent throughout the world in recent decades. In India, the country with the second biggest population, 7.5% of men suffer from OSA [49]. In the United States, an estimated 9% of middle-aged women and 24% of middle-aged men have at least mild OSA [54]. In Singapore, about 15% of the population is also estimated to be at risk [38]. With the spread of the obesity epidemic, the incidents of OSA will continue to rise.

Polysomnography (PSG), which monitors airflow, blood oxygen saturation, brain activity (EEG), heart rhythm (ECG), eye movements (EOG), and muscle activity (EMG), is the standard diagnostic test for OSA. However, it is complicated, expensive, and labor-intensive. Every PSG test attaches almost 20 sensors to the subject to monitor numerous body functions, and it costs around S$1000 and also requires professional technicians to stay an entire night to complete the diagnosis. In addition, the scarcity of PSG facilities results in severely limited accessibility and considerable waiting time. In Singapore, there are only two available sleeping laboratories, and the waiting time for a PSG is around three months. These limitations may result in the under-diagnosis and under-

treatment of millions of potential OSA patients. In fact, it is estimated that more than 80% of affected individuals remains undiagnosed [19].

Given the increasing prevalence of OSA and the limitations of PSG, researchers have investigated alternative diagnostic tools.

In medical field, questionnaires, such as the most famous Berlin Questionnaire [31, 11], the four-question STOP Questionnaire [10] and the clinical prediction model which are specifically derived for Singapore population in [27], were validated to be capable of predicting OSA. These questionnaires collect diagnostic information including age, gender, the occurrence and frequnency of waking up in the night, sleepiness in the daytime etc., and then the related information is analyzed to product a probability of being OSA patients. However, the answers of most questionnaires are objective and may require the assessment from patients' accompany. These two factors seriously affect the accuracy and feasibility of these medical prediction methods.

In computer science field, researchers explored the potential of OSA diagnosis using other modalities, such as nasal airway pressure [41], blood oxygen saturation [4], heart rate [40], and snore sound [3, 7, 16, 21, 26, 32, 33, 48, 52]. The first three modalities, though highly correlated with OSA diagnosis, require specific sensors to finish the collection. Specifically, nasal airway pressure needs to be measured by a sensor on the philtrum; blood oxygen saturation is usually measured by the pulse oximetry which should be placed on a thin part of patient's body; the measurement of heart rate requires the famous medical technique named electrocardiograph (ECG) which usually connects several sensors to the patient's body. These constrains obstruct them to be widely applied in OSA home screening.

Snoring, the earliest manifestation of upper airway abnormalities, is strongly associated with OSA, affecting 70% to 95% of OSA patients [42]. The snoring sound is generated by the vibration of soft tissues or the collapse of the upper airway due to air-

flow turbulence near a narrowed oropharynx [6]. Studies show that the upper airway of OSA patients has anatomical and functional abnormalities [5, 30]. As the upper airway acts as a variable acoustic filter in snore production, signs of abnormalities such as partial or total obstruction should be embedded in the snore sound. This hypothesis motivates researchers to detect useful features and patterns from snore-related signals to diagnose OSA. Moreover, compared with the three modalities above, the collection of snoring sound is much easier and cheaper without body contact and with minimal cost, which makes it ideal for OSA home screening. Currently, however, snoring sound is rarely used as a diagnostic criterion of OSA. Even for these few existing works, their investigation in the potential of using snoring sound to diagnose OSA is still far from satisfactory, in the aspect of both effectiveness and feasibility. Almost all these methods explore only the properties in snore episodes, the pure snores extracted from the whole-night snoring sound. Moreover, the results they provide cannot reflect the severity of OSA in a straightforward manner. Specifically, they just give a number indicating how probable the subject may be an OSA patient based on their defined measurement, but not tell him or her how sever the sleeping apnea is. Meanwhile, the useful information contained in silence episodes does not catch much attention in the research of OSA diagnosis using snoring sound. However, it is these silence episodes that are closely related to the clinical measurement of OSA.

OSA severity is clinically measured by the Apnea Hypopnea Index (AHI), which is defined as the number of respiratory events per sleeping hour. Respiratory events consist of two types: apnea event and hypopnea event. Apnea event refers to the complete cessation of nasal or oral airflow lasting for at least 10 seconds. Hypopnea event refers to a segment with 50% reduction of airflow for at least 10 seconds and is accompanied by decreased blood oxygen saturation. These respiratory events are usually reflected by special patterns in snore-related signals, especially for apnea events. For example,

one obvious phenomenon for apnea events is the occurrence of a long silence with no breath between two adjacent loud snore episodes, i.e. individual snores. Such particular patterns can assist the detection of respiratory events. Currently, the identification of respiratory events still requires time-consuming annotation manually done by professional technicians. However, almost all existing research focused on the abnormality of snore episodes while few studies delved into respiratory event detection. We are thus motivated to develop a system that performs better on OSA diagnosis by the automatic detection of respiratory events.

Home screening of OSA have also been investigated in recent years. Compared to PSG, these home-assisted devices are less expensive, less labor-intensive, less queuing time and easier to set up. Based on the rules of sleep apnea evaluation proposed by The Standards of Practice Committee of the American Sleep Disorders Association in 1994 [17], home screening OSA monitoring systems are categorized into four levels. Among these four levels, level I mainly refers to the traditional PSG; Level II, III and IV are portble OSA screening systems with different amount of sensors monitoring body functions. Specifically, Level II needs most sensors while Level IV requires fewest. Currently, systems in level II attracts the least interest from researchers because they still require complex measurements and are less user-friendly. Most existing portable monitors for OSA diagnosis belong to Level III, and systems in Level IV are also being explored such as those using snoring sound. For systems belonging to these three levels, they cannot identify sleep stages as PSG does, but they can detect respiratory events and measure the severity of OSA with AHI. The limitation of existing portable systems is that they did not fully investigate the latent useful information contained in collected signals and their performance and functionality still have much room for improvement. Therefore, we intend to implement a novel home screening system for OSA on mobile phone based on the proposed method .

1.2 Contributions

The main contributions of this project are as follows.

- A novel method is proposed to diagnose OSA with higher sensitivity and specificity compared with traditional diagnostic methods.
- Our method provides a reliably close approximation of the actual apnea event number. It has the potential to relieve technicians from the time-consuming annotations.
- A prototype named the Mobile Obstructive Sleep Apnea Diagnosis (MOSAD) has been developed on iPhone Operating System (iOS) based on the proposed method. It enables users to pre-diagnose OSA without attending PSG and makes home screening of OSA feasible.
- We are the first group to make a comparison among existing OSA diagnostic methods, not only validating their performance, but also improving upon their performance.

1.3 Organization

The body of this paper is organized as follows. Section 2 provides a comprehensive literature survey of the OSA diagnosis with snoring sound, the Conditional Random Field (CRF) and portable OSA diagnostic systems. Section 3 presents our proposed method to detect apnea events by using CRF. Validation and comparison experiments are presented in Section 4, and a prototype of the MOSAD implemented on iOS is

shown in Section 5. In Section 6, we summarize our work, draw conclusions and also suggest possible future research directions.

2 Literature Survey

2.1 OSA Diagnosis With Snore Sound Analysis

Most existing works diagnose OSA using information contained in snore episodes. The general framework of these methods are similar. Snoring sound is first segmented into individual components such as snore episodes, silence episodes, breathe episodes and speech. Then specific features are extracted from snore episodes. These features, containing the abnormality of OSA, are fed into classification models to diagnose OSA. Therefore, three main aspects are investigated in these existing works: segmentation of snoring sound, feature selection and model selection.

Abeyratne and Karunajeewa et al. contributed extensively to OSA diagnosis based on snore-related sound analyses. In their early research, they carried out pitch-jitter analysis to separate the signal into benign snore (BS), apnea snore (AS), and speech [1]. Benign snore was defined as a snore episode from healthy subjects while apnea snore represented snore episodes from OSA patients. Pitch-jitter analysis could classify snore episodes into AS class with 92.31% accuracy and BS class with 90.7% accuracy, suggesting that pitch might be a suitable candidate to identify apnea snores. Abeyratne and Karunajeewa et al. also designed an algorithm to segment snore-related-sound (SRS) into classes of pure breathing, silence, and voiced/unvoiced snores using pitch [3]. SRS was first classified into silence and non-silence based on log energy and number of zero crossing derived from the SRS, and then a pitch detector further classified the non-silence into breath and snore. To diagnose OSA, they proposed a novel feature, intra-snore-pitch-jump (ISPJ), which had a diagnostic sensitivity of 86% to 100% and a specificity of 50% to 80%. In 2007, they introduced a mixed-phase model to decompose the sleep signal into snore, breath, background noise, and speech signals [2]. With

this model, they proposed a general framework of source/total airway response (TAR) model to simulate the production of snore and breath. Through the analysis of the source signal and the TAR function, different signal components were extracted from SRS. Because TAR depicted the different structure of the upper airway during the production of apnea/benign snores, this source/TAR model also facilitated the classification of apnea snores and benign snores. In two of their recent papers in 2010 and 2011 [25][26], Abeyratne and Karunajeewa et al. investigated various parameters derived from pitch and TAR, for example, the mean and variance of pitch, center frequency, standard deviation of frequency, etc. These parameters were fed into a logistic regression model to estimate the probability of an OSA diagnosis with 89.3% sensitivity and 92.3% specificity. The performance of these pitch-related methods indicated that pitch could be used to diagnose OSA.

To help distinguish OSA patients from simple snorers, Sola-Soler and Jane et al. explored various features of snore episodes, such as pitch [43], snoring sound intensity [44], spectral envelope [46], and variability of snore parameters in time and frequency domains [47]. Moreover, they investigated the feasibility of applying a feedforward multilayer neural network to automatically detect snoring signal [23] and separate the simple snorers from the OSA patients [48]. In one of their recent papers in 2007, they claimed that subjects can be classified with a sensitivity higher than 93% and a specificity between 73% and 88% [48].

Cavusoglu and Ciloglu et al. [7] investigated the sequential properties of snoring episodes for OSA identification. Based on the recorded snoring signal, they derived a set of sequences, including snoring episode durations (SED), snoring episode separations (SES), and average snoring episode powers (SEP). To exclude the effects of slow variations in the baseline of these sequences, short time coefficient of variation (STCV) sequences, containing the coefficient of variation of the sample values in a "short" signal

frame, were investigated. Comparison experiments demonstrated that the statistical parameters obtained from the SED, SES, and the corresponding STCV sequences had the potential to distinguish simple snorers from OSA patients. However, the authors only revealed that those parameters were differentiable from simple snores and OSA patients by using the Student's t-test; its performance on a real data set was not examined.

Duckitt and Tuomi et al. [15] employed Hidden Markov Models (HMMs) to model different types of sounds by means of spectral-based features, including mel-frequency cepstrum coefficients (MFCCs), energy, and their first and second derivatives. HMM and MFCC were shown to be effective in speech recognition. Given the similarity of speech and snoring signal, this combination might be an appropriate method to isolate snoring sounds. Duckitt and Tuomi et al. claimed that their system was able to correctly identify snores with 82% to 89% accuracy. However, only six pieces of recording is used in this method to train and test the HMM model, which makes the result weak in demonstrating its effectiveness.

Ng and Koh et al., from the Nanyang Technological University in Singapore, cooperated with Abeyratne's group to further investigate the relation between snoring sound and OSA by detecting the difference of formant frequencies between benign snores and apnea snores [34, 33]. The first three formant frequencies were extracted from the LPC spectrum for analysis. They found that apnea snores exhibited higher formant frequencies than benign snores, especially the first formant frequency (F1). They reported that the optimal threshold value of F1 that differentiates apnea snorers from benign snorers is 470 Hz in one paper [34], but claimed it to be 720 Hz in another paper [33]. Therefore, the optimal threshold for F1 may need further investigation. In 2009, this group proposed to use a nonlinear mode, wavelet bicoherence (WBC), to process snore signals and diagnose OSA [32]. They defined two novel markers, peak frequency component at F1 (PF1) and peak sum frequency (PSF), to differentiate apnea and benign snores.

The result showed that the nonlinear mode interactions in apnea snores were less self-coupled and usually occupied higher and wider frequency ranges than those in benign snores. The sensitivity and specificity values, which were both between 85.0% and 90.7%, indicated a promising prospect on nonlinear dynamic analysis of snore signals. In that paper, they also explored the relation between AHI and the proposed markers (PF1 and PSF), which likely took the functional form of exponential or power. This was the first paper that investigated the relation between AHI and diagnostic parameters.

Although the aforementioned methodologies had high accuracy, some of them were just validated on the classification of benign snores and apnea snores. The performance of the other methods was shown to be promising, however, they only had the ability to roughly classify subjects into the OSA group and the healthy group because the results they obtained are not directly related to OSA diagnosis justification, specifically apnea and hypopnea events.

In [22], Hou and Xie et al. defined a respiratory event as an interval longer than 10 seconds between two adjacent snore events. They attempted to detect respiratory events using a dynamic threshold for endpoint detection (EPD) of snore episodes. Although their target parameter was directly related to the calculation of AHI, their definition and method had several weaknesses. First, the pattern they defined for respiratory events is less descriptive for capturing hypopnea events that are not strictly associated with absolute silence episodes. Second, EPD tends to incorrectly segment noise as snore and miss any soft breaths occurring within long silence episodes. Moreover, no experiments were conducted to validate if the detected events were real respiratory events. In other words, those detected respiratory events might not be the real apnea or hypopnea events even if the calculated AHI was close to the one diagnosed with PSG. The results would have been more strongly supported if the detected events had been compared with those annotated by technicians from a sleeping laboratory.

2.2 Conditional Random Field

Developed by Lafferty et al. [29] in 2001, CRF is defined as a technique that, given G, a graphical structure describing the clique template for each instance, and an observation sequence $\vec{x} = (x_1, x_2, ..., x_n)$, obtains the best output label sequence $\vec{y} = (y_1, y_2, ..., y_n)$ by optimizing the conditional distribution $\Pr[\vec{y} \mid \vec{x}]$. CRF is an effective probabilistic model for sequential labeling and has been widely adopted in various domains such as part-of-speech tagging [29] and text segmentation [28] in natural language processing, image labeling [20] and object recognition [50] in computer vision, gene prediction in bioinformatics [13], etc.

Another well-known model for sequential structures is the Hidden Markov Model (HMM). However, the strong independence assumption HMM makes between the observation variables attenuates the accuracy of systems derived from it. In comparison, CRF performs better as it does not need to model the dependencies among observation variables. Moreover, while the application of HMM is limited to linear sequential structures, CRF can be generalized to arbitrary structures and can better capture the dependencies among sequential variables. Thus, CRF has a clear advantage for learning the relations of sequential acoustic components, esp. those found in snoring sound.

During an apnea event, patients often emit a clogged snore during inhalation, not breathe for a long period, generate abnormal sounds as they struggle to breathe, and then produce a sudden and loud snore to complete the respiratory cycle. This pattern of sound involves not only the long silence but also the snore episodes directly before and after. The dependencies between these snore and silence episodes make sequential labeling of respiratory events a valid and potentially highly effective approach to OSA diagnosis.

However, a manually annotated respiratory event does not strictly associate with an

individual snore episode or silence episode. In fact, one manually annotated respiratory event may include several snore, breath, and silence episodes, and its boundaries are inconsistent with those demarcated by automatic segmentation. As observed, one dominant, long silence episode usually occurs inside every apnea event and some hypopnea events. Given this particular pattern, the problem of respiratory event annotation thus transforms into that of sequentially labeling these silence episodes, a problem which CRF is well suited to solve.

Therefore, in this thesis, we propose a relational learning diagnostic method using CRF to identify apnea and hypopnea events. Features extracted from snore and silence episodes are fed into a CRF model as observations. Manually annotated respiratory events are associated with specific silence episodes and these silence episodes are then used as the output of CRF model. Based on the observations and the output, a CRF model is trained to label respiratory events.

2.3 Portable OSA Diagnosis System

The Standards of Practice Committee of the American Sleep Disorders Association proposed four levels of studies on sleep apnea evaluation in 1994 [17]. Among these four levels, portable monitoring is possible for Level II, III and IV. Table 2.1 provides a comparison among these three levels. Level II attracts the least interest from researchers because it still requires complex measurements and is less user-friendly. Most existing portable monitors for OSA diagnosis belong to Level III, but systems in Level IV are also being explored. Although facilities in these two levels cannot identify sleep stages as PSG does, they can detect respiratory events and measure the severity of OSA with AHI.

In recent years, a number of home-assisted OSA diagnosis systems have been developed [35, 14, 37]. Compared to PSG, these home-assisted devices are less expensive,

	Level II	Level III	Level IV
Description	Unattended PSG	Modified portable sleep apnea testing	Continuous single or dual bioparameter recording
Measures	Minimum of 7, including EEG, EOG, chin EMG, ECG or heart rate, airflow, respiratory effort, oxygen saturation	Minimum of 4, including ventilation (at least two channels of respiratory movement, or respiratory movement and airflow), heart rate or ECG, oxygen saturation	Minimum of 1: oxygen saturation, airflow, or chest movement

Table 2.1: Comparison among three levels of portable OSA monitoring

less labor-intensive, and more convenient to set up. However, their performance and functionality are still far from satisfactory.

Snoring sound can be collected without body contact and additional cost. Therefore, it has attracted considerable interests. Five existing systems utilizing snoring sound are examined here as representatives of portable OSA diagnosis systems: ARES [51], CID102L8 [37], Stardust II [53], MORFEAS [14], and Ashida's system (based on sound and SpO2 monitoring) [35]. Table 2.2 presents comparisons among these systems in five aspects, namely, the level they belongs to, the channels they collect data from, the hardware they use, whether the questionnaire is included and whether automatic diagnosis result can be given without technicians attending.

Even though the simplest system of the five, MORFFEAS, merely records snore-related sounds and transmits the data to technicians for diagnosis. Nonetheless, it is quite a complex system with several different modules, including a recorder, a memory storage unit, and a networking unit to transmit data to a sleep laboratory. The most significant shortcoming of this system is that technicians are still required to attend the diagnosis, which makes it unsuitable for OSA home screening. The other four systems not only utilize SRS but include other modalities such as tracheal breath sounds, nasal

	ARES	CID102L8	Stardust II	MORFEAS	Ashida's System
Level	III	III	III	IV	IV
Channels	snoring level, arterial oxygen saturation, pulse rate, head movement	tracheal breath, nasal flow, body position, arterial oxygen saturation, heart rate, etc.	snoring sound, oronasal airflow, arterial oxygen saturation, pulse rate, body position, etc.	snore sound	snore sound, SpO2
Hardware	a brain monitor affixed to forehead	recorder, analyzer	Stardust II device	recorder, networking unit, memory unit	an IC recorder, a simple SpO2 monitor
Questionnaire	Yes	No	Yes	No	No
Automatic diagnosis	Yes	Yes	Yes	No	Yes

Table 2.2: Comparison of five portable OSA diagnosis systems

flow, thoracic and abdominal movements, body position, and arterial oxygen saturation. Although, these additional channels may improve diagnostic accuracy, they require sensors with body contact, have complicated setups, and are even less user friendly. More importantly, all five systems utilize only a few simple features of the snoring sound, leaving the rich diagnostic information in snoring sound unexplored. An ideal system for OSA home screening should have no body contact with subjects, be easy to set up, require no additional hardware and no technicians attending, and produce reasonable diagnostic accuracy. Given these factors, we proposed a prototype named MOSAD, which is merely a software on iOS, to diagnose OSA using only recorded snoring sound.

3 Apnea Events Detection using CRF

Figure 3.1 presents a schematic of the proposed system, which consists of three major components:

- Snoring sound segmentation. This part identifies the acoustic signal components: silence, snore and non-snore episodes. Apnea and hypopnea events annotated by technicians are then associated with detected silence episodes to generate an event label.
- CRF observation extraction. In this part, observations for CRF are extracted from the components above, which, together with their event labels, serve as the training data for CRF.
- Apnea events detection. A CRF model is trained to detect apnea events for OSA diagnosis.

3.1 Data Collection

We implemented an iPod Touch software to record snoring sound overnight. The iPod Touch was placed on a desk beside the bed, and the distance from the microphone to the mouth of the patients was approximately 50 cm. All of the recorded snoring signals were temporarily stored in the iPod Touch and transferred to computers later for processing. A total of 28 pieces of recording were collected during routine PSGs in the sleeping laboratory of the National University Hospital in Singapore. Corresponding

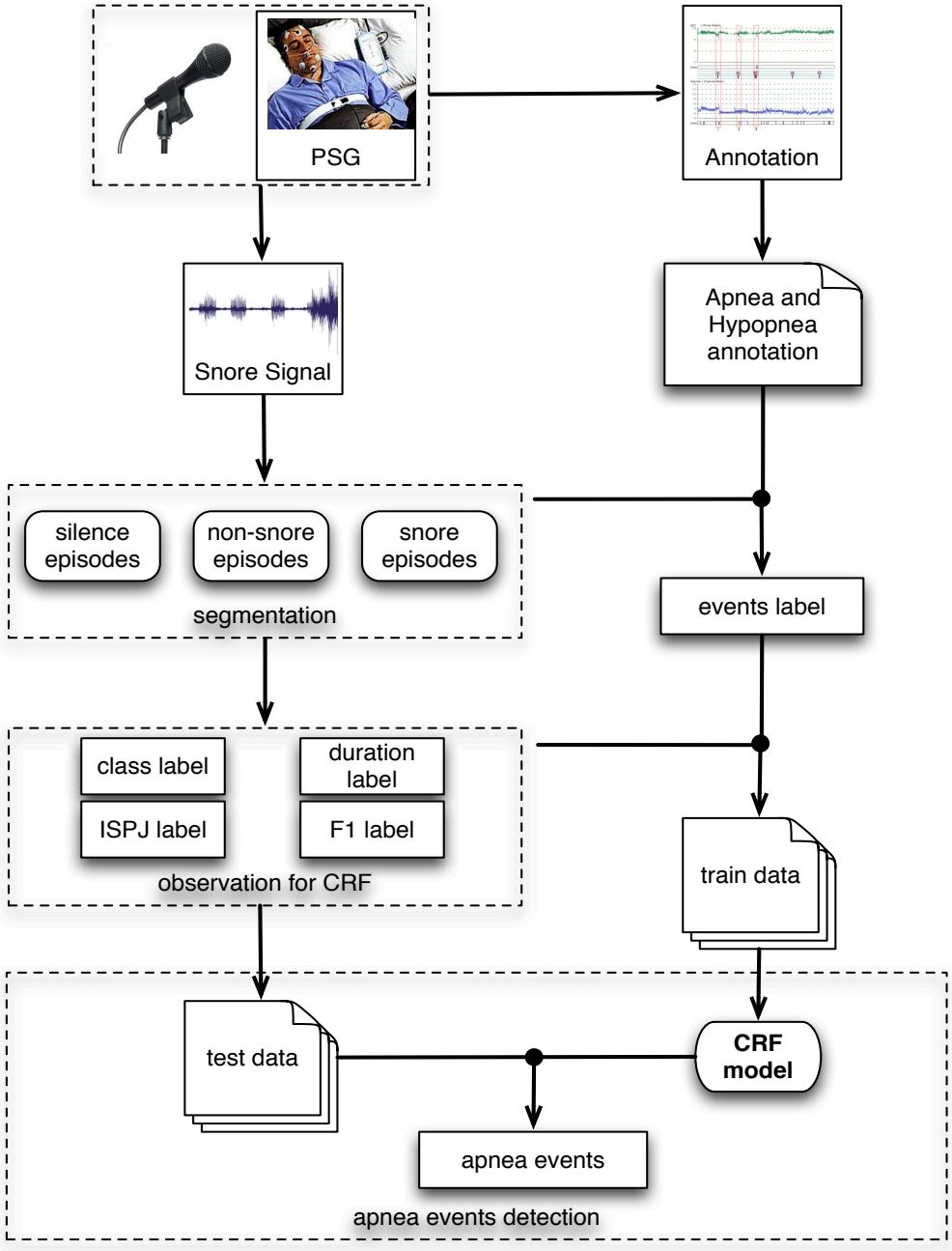


Figure 3.1: Flowchart of OSA diagnosis system with CRF

PSG reports for all 28 subjects and respiratory events annotation done by technicians for 14 subjects were also collected (annotations for the rest subjects are not successfully collected due to mis-operations or early deletions). All samples were collected with sampling rate of 44.1 kHz and quantizing precision of 16 bits and were stored in wave files.

3.2 Automatic Segmentation

Energy and zero crossing rate (ZCR) are features frequently used to detect the boundaries of sound episodes. Let N be the length of a frame and $x_k[i]$ be the ith sample of the kth frame. The energy of the kth frame is calculated as

$$E_k = \sum_{i=0}^{N-1} x_k[i]^2.$$

Following the method introduced in [8], we derive thresholds for energy and ZCR and label a frame as a sound frame if its energy and ZCR are above the thresholds. The threshold of energy is computed as

$$T_e = \min\{I_1, I_2\}$$

where

$$I_1 = a \times \left[\max_k \{E_k\} - \min_k \{E_k\}\right] + \min_k \{E_k\},$$

$$I_2 = b \times \min_k \{E_k\}.$$

The mean ZCR is calculated from the training data set, and the ZCR threshold is computed with the mean ZCR as follows:

$$T_z = c \times mean(ZCR).$$

Parameters a, b and c are the same as those used in [8] (a = 0.02, b = 3, c = 0.3).

After labeling, frames are merged into silence or sound episodes in two steps. First, continuous frames with the same label are merged. Second, small fragments are merged based on a threshold β, which is experimentally set to 2. More specifically, if two episodes with the same label are longer than β frames and are separated by another episode shorter than β frames, then they are merged into one episode. This smoothing process segments the snoring sound into silence episodes and sound episodes.

The next stage is to identify snore episodes from previously detected sound episodes. As defined in [3], *a snore episode is a breath record with at least one portion of it containing sound with a detectable pitch.* Pitch can be detected by applying the YIN algorithm [12] on each of the frame in a sound episode. If there is at least one frame with a detectable pitch, this sound episode is recognized as a snore episode; otherwise, it is labeled as a non-snore episode.

Having segmented and labeled the snore-related signal into silence, snore, and non-snore episodes, the relation between them will be investigated with CRF.

3.3 Apnea Event Detection using CRF

3.3.1 CRF Briefing

In the proposed method, CRF is used as a sequential labeling tool. Therefore, only the basic idea of CRF is introduced in this section. Detailed explanation can be found in [28, 29].

Clique, a graph C generated for each instance during training and inference, repre-

sents the relation that CRF is supposed to learn. The portion in observation sequence $\vec{x}$ and output sequence $\vec{y}$ belonging to C are represented with $\vec{x_C}$ and $\vec{y_C}$, respectively. Depending on the defined combination of $\vec{x_C}$ and $\vec{y_C}$, a potential function ψ_C is generated to represent the joint probability of $\vec{x_C}$ and $\vec{y_C}$, i.e. $\psi_C(\vec{x_C},\vec{y_C})$. However, because the potential function can be an arbitrary function, it should be normalized into a proper probability measure. Therefore, the joint probability of the observation and output sequence is

$$\Pr[\vec{x},\vec{y}] = \frac{1}{Z}\prod_{C\in\mathbb{C}} \psi_C(\vec{x_C},\vec{y_C}),$$

in which Z normalizes the potential function and is computed as

$$Z = \sum_{(\vec{x},\vec{y})}\prod_{C\in\mathbb{C}} \psi_C(\vec{x_C},\vec{y_C}).$$

$\mathbb{C}$ is the set of all cliques in the whole sequence. Thus, the conditional probability is obtained as

$$\begin{aligned}
\Pr[\vec{y} \mid \vec{x}] &= \frac{\Pr[\vec{x},\vec{y}]}{\Pr[\vec{x}]} \\
&= \frac{\Pr[\vec{x},\vec{y}]}{\sum_{\vec{y}'}\Pr[\vec{y}',\vec{x}]} \\
&= \frac{\frac{1}{Z}\prod_{C\in\mathbb{C}}\psi_C(\vec{x_C},\vec{y_C})}{\frac{1}{Z}\sum_{\vec{y}'}\prod_{C\in\mathbb{C}}\psi_C(\vec{x_C},\vec{y_C}')} \\
&= \frac{1}{Z(\vec{x})}\prod_{C\in\mathbb{C}}\psi_C(\vec{x_C},\vec{y_C})
\end{aligned}$$

where

$$Z(\vec{x}) = \sum_{\vec{y}'}\prod_{C\in\mathbb{C}}\psi_C(\vec{x_C},\vec{y_C}')$$

is also a normalization denominator.

Therefore, given a sequence of observations $\vec{x}$, CRF calculates the conditional probability $\Pr[\vec{y} \mid \vec{x}]$ for each possible output sequence $\vec{y}$, and selects the one with the maximal conditional probability as the output label. In fact, the most possible output sequence for an observation sequence can be efficiently determined using the famous Veterbi algorithm [39].

3.3.2 Association from respiratory events to silence episodes

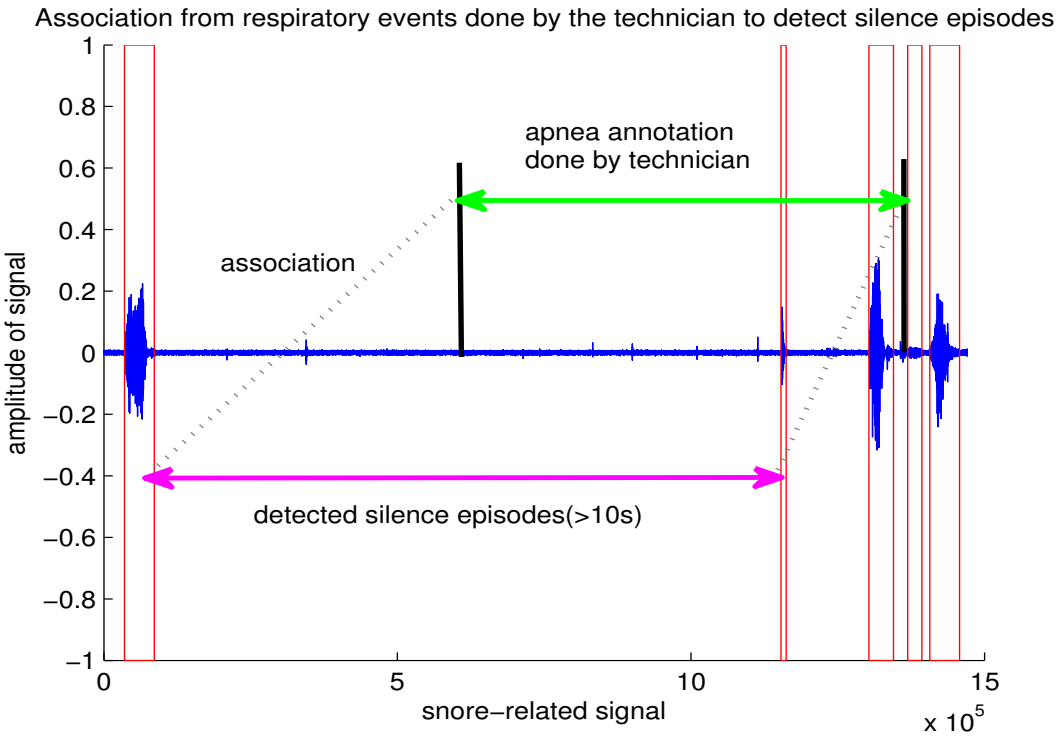


Figure 3.2: Association from respiratory annotations to silence episodes

To detect respiratory events with CRF, the output of our CRF model should somehow be related to the real respiratory events. Technicians in the sleeping laboratory annotate apnea and hypopnea events based on multi-channels. These annotations serve as the real respiratory events. These manually annotated respiratory events have inconsistent boundaries with the acoustic signal components detected by our segmentation method.

However, we observe that there is always a relatively long silence episode in the annotated respiratory events, especially in apnea events. We call these silence episodes apnea silences or hypopnea silences. According to the definition of apnea events, it is also the silence episodes longer than 10 seconds that most strongly associate with apnea events. Therefore, in this paper, we associate a real apnea and hypopnea event to the longest silence episode between its start and end positions as shown in Figure 3.2. The associated silence episodes are labeled as apnea or hypopnea according to the type of respiratory event (1 for apnea event and 2 for hypopnea event) while the other episodes are all labeled as 0. These event labels act as ground truth for CRF training.

As shown in Figure 3.2, short, softer snores often occur before the beginning or after the end of apnea silence and hypopnea silence, but these snore sounds are not the dominant snore resulting in apnea events. We call these snore episodes abnormal snores. Based on our observation, these abnormal snores are usually less than one second. Therefore, we define dominant snore as the first snore that are longer than one second surrounding the silence episode. The sequence from the apnea silence to the dominant snores may help to recognize specific patterns for apnea and hypopnea events.

3.3.3 Clique for CRF

In our CRF model, a clique is defined to be an undirected graph with seven episodes as shown in Figure 3.3. In this figure, x represents the observation vector of each episode while y represents the label generated by CRF. Since our CRF model only labels silence episode to be either kind of respiratory event, in the following description, we will just discuss the clique generated for silence episodes.

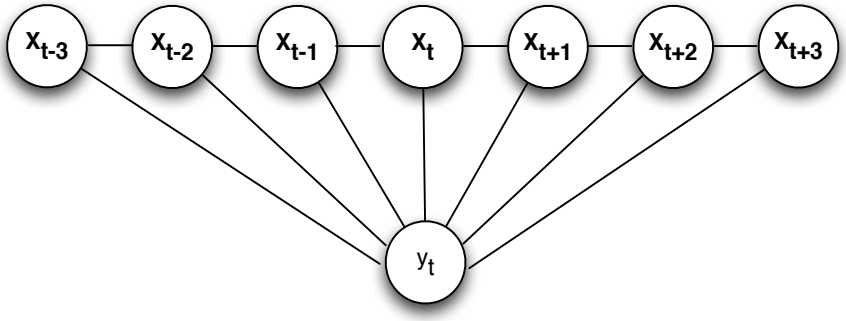


Figure 3.3: Clique for CRF

3.3.4 Observation Extraction for CRF Training and Testing

To diagnose OSA, we propose a relational learning method using CRF to detect respiratory events. As CRF requires discrete training data, we extract four features from each episode and label them using discrete values, which serve as the observations of CRF.

The first observation is the class label. After automatic segmentation into episodes with the method introduced above, each episode is associated with a class label (silence = 1, breath = 2 and snore = 3). However, Respiratory events are mainly reflected by the relation between silence episodes and snore episodes. From now on, only silence episodes and snore episodes are considered.

The second observation is the duration label. The duration of silence episodes is the most important criterion for finding apnea events, which can be indicated from the definition of respiratory events. Table 3.1 shows the duration label assignments .

The two categories for snore episodes ($t \leq 1$ and $t > 1$) are defined to differentiate the dominant snore episodes from those abnormal snores. The category of $0 \leq t \leq 3$ for silence episodes is defined to represent silence episodes between dominant snore episodes and abnormal snores. These three categories together depict the association of silence episode to its surrounding dominant snore episodes.

While all hypopnea events contain silences, most events are associated with silence

silence		snore	
duration (seconds)	label	duration (seconds)	label
$0 \leq t \leq 3$	0	$t \leq 1$	-1
$3 < t < 5$	1	$t > 1$	-2
$5 \leq t < 10$	2	-	-
$10 \leq t < 60$	3	-	-
$60 \leq t \leq 120$	4	-	-
$t > 120$	5	-	-

Table 3.1: Duration label for silence and snore episodes

episodes less than 10 secs. Therefore, we define the category of $5 \leq t < 10$ to facilitate the detection of hypopnea events.

The two categories of $10 \leq t < 60$ and $60 \leq t \leq 120$ are defined to detect the silence episode within an apnea event. The lower bound 10 seconds is determined according to the definition of apnea events, whereas the upper bound 120 seconds is set based on the longest duration of apnea events (mean: 55.57 seconds; standard derivation: 21.95 seconds; range: 16.9 - 102.4 seconds) in our data set.

The range larger than 120 seconds represents episodes of the normal, silent unobstructed breathing, and the category in $3 < t < 5$ bears no special meaning.

The last two observations are defined to characterize abnormalities of snore episodes. Based on the snore source/upper airway model proposed in [2], snoring sound $f(n)$ can be decomposed into the convolution of vocal excitation $s(n)$ and vocal tract $h(n)$:

$$f(n) = s(n) \circ h(n).$$

Vocal excitation, the original sound or vibration, can be described in terms of pitch, and the vocal tract, a filter that shapes the raw vibration when it gets passed, can be depicted in terms of formants. In the production of apnea and hypopnea events, not only the vocal excitation, but also the vocal tract of the subjects are abnormal. When

snore source passes the totally or partially obstructed upper airway, the transfer of snore sound energy at specific frequency bands is attenuated while the maximal energy at the resonance frequencies are allowed to pass. This phenomenon can be captured using pitch and formant frequencies.

Pitch is defined as the fundamental frequency of snoring sound vibration. In previous studies [1, 3, 25, 26, 48, 45], pitch showed potential as a candidate for OSA diagnosis. As introduced in the literature survey, the authors in [3] found that pitch jump can be detected in some snore episodes from OSA patients, and they defined a new feature, ISPJ, to capture this abnormality. A snore episode is labeled as an ISPJ snore if there are at least q frames in the snore episode with pitch period higher than a threshold γ. Then the percentage of ISPJ snores in all snore episodes is used to classify subjects into the OSA group and the healthy group. In the proposed method, a binary value named ISPJ label is assigned to each snore episode. This value is the third observation that describes the abnormal characteristics in snore episodes. The ISPJ label for a silence episode is always set to zero.

Formant frequencies represent the resonance pattern of the upper airway. The first three formant frequencies, especially the first one (F1), have been shown to carry useful information for OSA diagnosis [33, 34]. Therefore, for the last observation, we introduce an F1 label based on the F1 values as calculated by a 12th-order Linear Prediction Coefficient (LPC) analysis. If the F1 is greater than the pre-set threshold f, the label is set to one; otherwise, it is set to zero. The F1 label for silence episodes is always set to zero.

With these four observations, each episode is represented as a four-dimension vector

$$< ClassLabel, DurationLabel, ISPJLabel, F1Label >$$

Vectors for all episodes form the test data for CRF. Combined with the corresponding

respiratory events labels, they are used to train the CRF model.

3.3.5 Observation conjunction

Sequential acoustic components are conjuncted to depict special patterns that may associate with respiratory events. Since only the labeling of silence episodes is concerned, we merely list the conjunction of observations for the clique generated for silence episode in Table 3.2. In this table, *E* is the abbreviation for episode.

Conjunction	Description
E_0	current silence
E_{-1}	the previous snore
E_1	the following snore
E_{-3}	the snore before E_{-1}
E_3	the snore following E_1
$E_{-1}E_0$	current silence and the previous snore
E_0E_1	current silence and the following snore
$E_{-1}E_0E_1$	current silence, the previous and the following snore
$E_{-3}E_{-2}E_{-1}E_0$	current silence and the previous two snores
$E_0E_1E_2E_3$	current silence and the following two snores
$E_{-3}E_{-2}E_{-1}E_0E_1$	current silence, the following snore, the previous two snores and silence in between
$E_{-1}E_0E_1E_2E_3$	current silence, the previous snore, the following two snores and silence in between
$E_{-3}E_{-2}E_{-1}E_0E_1E_2E_3$	current silence, the previous two snores, the following two snores, and silences in between

Table 3.2: observation conjunction used in CRF model

In this table, the five first-order conjunctions represent the pattern of the episode itself. The following two second-order conjunctions are defined to capture the relation between the current silence episode and its previous or following snore episode separately while the third-order conjunction reflects these two kinds of relations at the same time. The second-order and third-order conjunctions are defined to represent the case

that the previous and/or following snore episode are the dominant snore episodes that cause apnea or hypopnea events. The two fourth-order conjunctions and fifth-order conjunctions mean the previous or the next snore episode of the current silence episode may not be the dominant snore episode and the current silence episode may need to be associated with the second previous or following snore episode. The final seventh-order conjunction represents the case that both the previous and following episodes are not the dominant snore episodes and the relation between current silence episode and further snore episodes in both sides should be learned.

As a summary, in this section, we introduced the process of data collection, the segmentation of snoring sound and the detection of apnea events utilizing CRF.

4 Experiments

We conduct experiments to:

- decide the mean ZCR and the best parameter setting for F1 threshold,
- train the CRF model,
- show the performance of our method on OSA diagnosis,
- show the performance of our method on the detection of apnea events,
- compare the performance of our method with two snore-episode-related methods,
- compare the performance of our method with a respiratory-event-related method.

4.1 Experimental Parameters

Unless specifically illustrated, the parameter set in Table 4.1 is used for all experiments in this thesis.

Group	Parameter	Value
Data	sampling rate	44100 sample/s
	bit quantization	16 bits
Segmentation	window size	100 ms
	overlap	50 ms
	β	2
ISPJ label extraction	q	1
	γ	19 ms
OSA diagnosis	AHI cutoff	15

Table 4.1: Parameter setting for experiments

The open source implementation named CRF++ [1] is adopted in the experiments for CRF training and testing.

[1] http://crfpp.googlecode.com/svn/trunk/doc/index.html

4.2 Training Process

In this step, two thresholds are trained using a set of manually annotated snore episodes: mean ZCR, used for segmentation, and F1 threshold, used to generate F1 labels for CRF. The training process of CRF will also be explained in this section.

4.2.1 Parameter Determination

Two hundred snore episodes from each of seven subjects (BMI: 30.83 $\pm$ 7.93 kg/m^2; AHI: 33.44 $\pm$ 28.01 events/h, ranging from 6.1 to 90.4) are manually annotated and extracted. These 1400 snore episodes yield a mean ZCR of 0.0922, which is then used as the ZCR threshold.

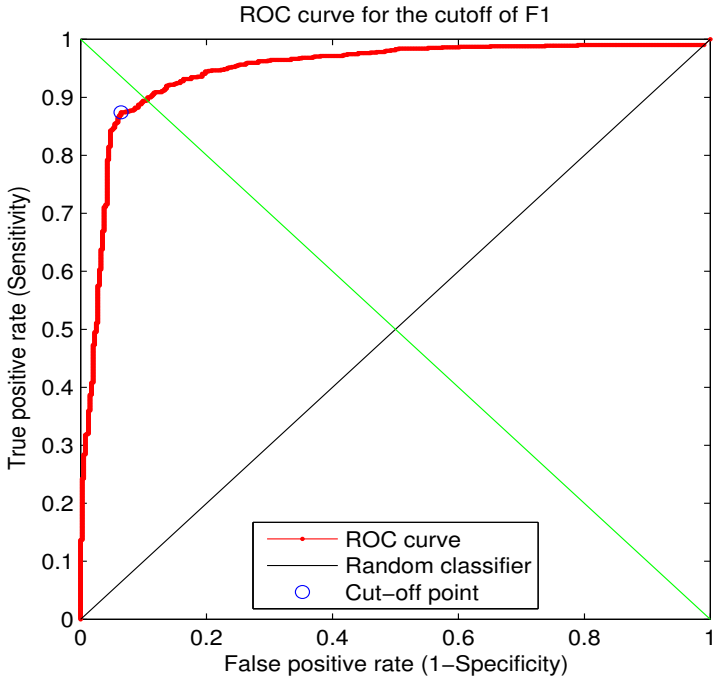


Figure 4.1: ROC analysis for the threshold of F1

Two papers [33, 34] by Ng and Koh et al. report two different thresholds of F1, 470 Hz and 720 Hz, as the optimal threshold to differentiate between benign and apnea

snores. The discrepancy of these two thresholds is too large for us to decide which one to use. Therefore, we independently derived the F1 threshold based on our own data set, with AHI = 15 as the cutoff for benign snores versus apnea snores. F1 is determined by using a 12th-order LPC analysis for each snore episode. A receiver operator curve (ROC) analysis is then applied to obtain the best cutoff. As shown in Figure 4.1, the optimal threshold for F1 is determined to be 451.4991 (sensitivity: 87.43%, specificity: 93.07%, AUC: 0.94551).

4.2.2 Training for CRF Model

Due to misoperation, only 14 of the 28 pieces of overnight recordings have manually annotated respiratory events. Of those, 10 pieces (BMI: 28.71 $\pm$ 6.03 kg/m^2; AHI: 42.92 $\pm$ 30 events/h, ranging from 1.9 to 94.7; total sleeping time: 327 $\pm$ 58.26 seconds) are chosen as training data for CRF. Table 4.2 lists the related statistics of CRF training data, including BMI, AHI, number of apnea events (AEN) and hypopnea events (HEN), and total sleeping time (TST).

Training data	BMI	AHI	Number of Events		TST (seconds)
			AEN	HEN	
S1	33.20	33.6	100	107	369.5
S2	31.64	75	356	49	324.0
S3	30.07	69.6	260	152	355.0
S4	36.89	94.7	43	391	275.0
S5	22.99	40.8	174	107	413.0
S6	30.54	18.4	3	62	211.5
S7	20.06	14.5	15	57	298.5
S8	35.18	55.9	44	260	326.5
S9	26.27	1.9	1	11	386.5
S10	20.28	24.8	85	45	314.5

Table 4.2: Statistics of subjects used for CRF training

For each piece of recording, we create training files in which each row represents <class label, duration label, ISPJ label, F1 label, event label>. These 10 files are then concatenated into one to be fed into the CRF model for training.

4.3 Testing with CRF

The remaining 18 pieces of recording (BMI: 31.11$\pm$ 5.91 kg/m^2; AHI: 32.78 $\pm$ 32.30 events/h, ranging from 0.5 to 98.6; total sleeping time: 332.42 $\pm$ 43.22 seconds) are used as the testing data set. Related statistics are collected from the PSG report and listed in Table 4.3.

Testing data	BMI	AHI	Number of Events		TST (seconds)
			AEN	HEN	
S11	40.94	86.1	392	22	288.5
S12	32.86	90.4	274	199	314.0
S13	30.17	43.3	135	85	304.5
S14	38.48	68.8	358	16	327.0
S15	38.52	98.6	307	316	379.0
S16	34.19	27.8	25	69	203.0
S17	36.49	46.7	135	157	375.5
S18	33.52	0.5	0	3	353.5
S19	22.31	22	11	113	338.5
S20	31.06	12.6	18	46	305.0
S21	25.28	40.2	51	169	352.0
S22	37.11	72	355	33	323.5
S23	25.35	15.7	36	57	355.0
S24	29.74	57.3	165	189	371.0
S25	21.64	6.1	4	30	333.5
S26	28.62	10.3	45	8	309.5
S27	30.30	15.1	65	28	373.0
S28	23.36	2.5	7	9	377.5

Table 4.3: Statistics of subjects used for CRF testing

Among the 18 pieces of data shown in Table 4.3, only the first four (S11–S14) come with respiratory event annotations. These annotations are associated with silence episodes using the method introduced in Section 3.3.3. The associated silence episodes serve as a reference to evaluate the consistency between the respiratory events detected by CRF and those reported by PSG.

Due to a technical failure, the respiratory event annotations of the remaining 14 pieces (S15–S28) have not been collected. However, we do have their PSG reports, which clearly describe the apnea and hypopnea numbers, AHI, total sleeping time, and the longest duration of respiratory events. To assess the effectiveness of apnea event detection using CRF, the number of detected apnea events and the apnea index (AI) value are compared with those reported by PSG.

4.4 Performance of Respiratory Event Detection

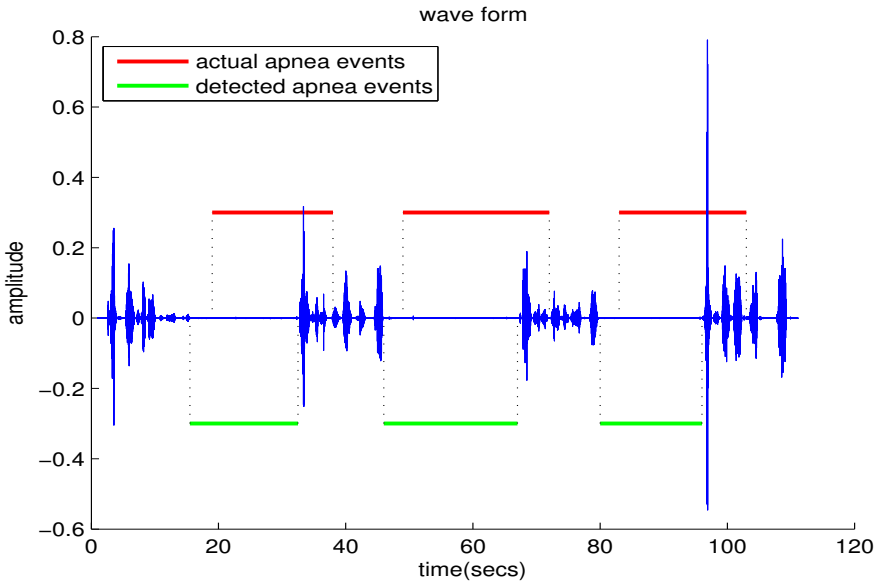


Figure 4.2: Effect of CRF on apnea event detection

Figure 4.2 shows the effect of apnea event detection using CRF on a two-minute snoring sound clip. As shown in this figure, manually annotated apnea events are always behind the detected ones. This is because technicians, when labeling apnea events, mainly refer to arterial oxygen saturation level and highlight sustained decrease of saturation, which appears a few seconds later than the cessation of breathing. Although the time differences exist, the detected apnea silences indeed reflect the occurrence of apnea events.

For convenience, we abbreviate the results obtained by different methods for the following experiments using the format "$\text{Result}_{\text{method}}$" (e.g. AEN_{PSG}, P_{ISPJ}, AI_{CRF} and AHI_{EPD}).

Subject	AEN_{PSG}	AEN_{CRF}	HEN_{PSG}	HEN_{CRF}
S11	392	325	22	14
S12	274	237	199	5
S13	135	107	85	6
S14	358	388	16	16
S15	307	295	316	12
S16	25	27	69	3
S17	135	101	157	14
S18	0	5	3	0
S19	11	11	113	1
S20	18	24	46	1
S21	51	32	169	4
S22	355	237	33	10
S23	36	42	57	5
S24	165	160	189	12
S25	4	2	30	0
S26	45	46	8	1
S27	65	70	28	0
S28	7	10	9	2

Table 4.4: Comparison of AEN_{PSG} and AEN_{CRF} and that of HEN_{PSG} and HEN_{CRF}

Table 4.4 compares the number of respiratory events reported by PSG and that detected by CRF. As we can see in the table and Figure 4.3, the number given by CRF

approximates that given by PSG in most cases. However, for subjects with numerous apnea events, the detected number slightly deviates from the actual one. An apnea event may not be detected because its apnea silence is split into two shorter silence episodes by a mislabeled sound episode. For patients with fewer apnea events, the detected number is slightly higher than the one reported by PSG, possibly because silence episodes that are normal or associated with hypopnea events are erroneously labeled as apnea silences.

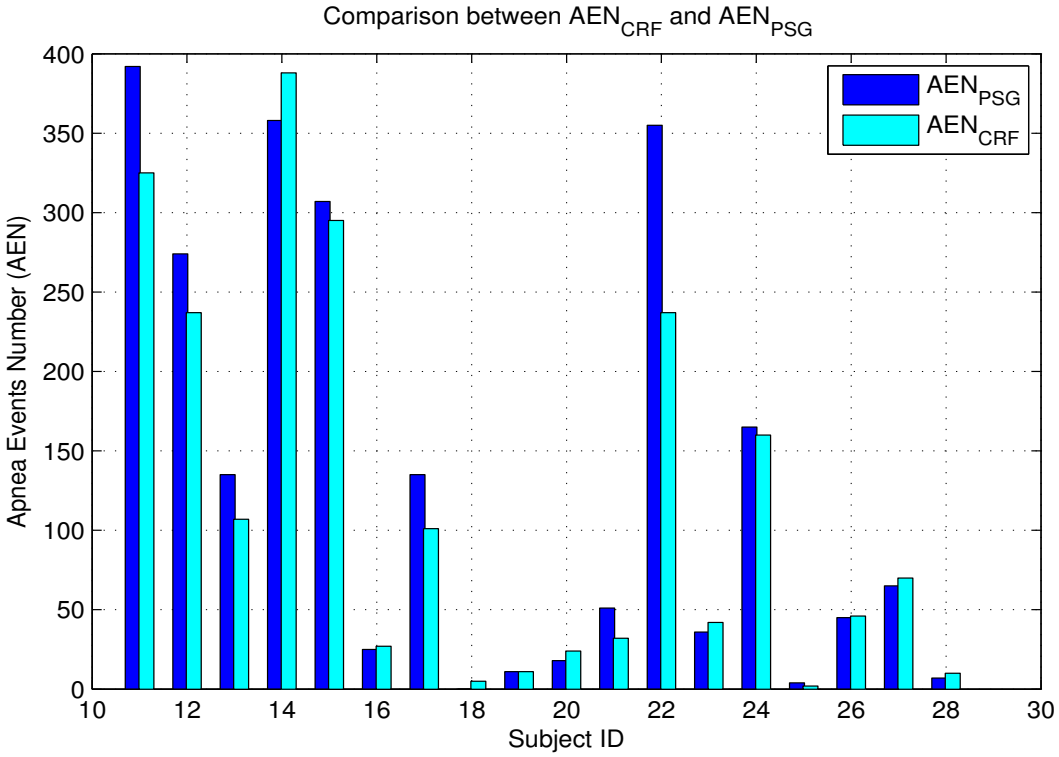


Figure 4.3: Comparison between AEN_{PSG} and AEN_{CRF}

Table 4.4 also illustrates that hypopnea events are not effectively recognized using CRF. This may be due to the differences between the definition of hypopnea event and our criteria for associating respiratory events to silence episodes. Hypopnea event is defined as a segment with a 50% amplitude reduction of airflow for at least 10 seconds, accompanied by a decrease of blood oxygen saturation. This definition implies that

there may still be sounds within a hypopnea event, which agrees with our observation that many hypopnea events cannot be associated with silence episodes longer than 10 seconds. Moreover, since apnea and hypopnea events are both associated with respiratory cessation, hypopnea events with silence episodes that are longer than 10 seconds may be incorrectly identified as apnea events.

AHI, which is the sum of apnea index (AI) and hypopnea index (HI), measures the severity of OSA. They are respectively computed as follows:

$$AHI = \frac{AEN + HEN}{TST},$$

$$AI = \frac{AEN}{TST},$$

and

$$HI = \frac{HEN}{TST}.$$

AI, the number of apnea events per sleeping hour, is originally the definitive indicator of OSA severity [19]. Although AI is currently replaced with AHI, it is still indicative of apnea events, the hallmark of OSA. As shown in Figure 4.4, the AIs obtained by CRF approximate those reported by PSG well. For subjects with high reported AIs (e.g. S11 and S22), the detected AIs carry more deviation. The difference, however, does not affect the categorization of these subjects.

	CRF
Optimal threshold	AI = 5
TP FP FN TN	12 1 1 4
Sensitivity	92.31%
Specificity	80.00%

Table 4.5: Performance of OSA diagnosis using CRF

In our method, we use AI = 5, the threshold adopted for OSA diagnosis in [19]. As

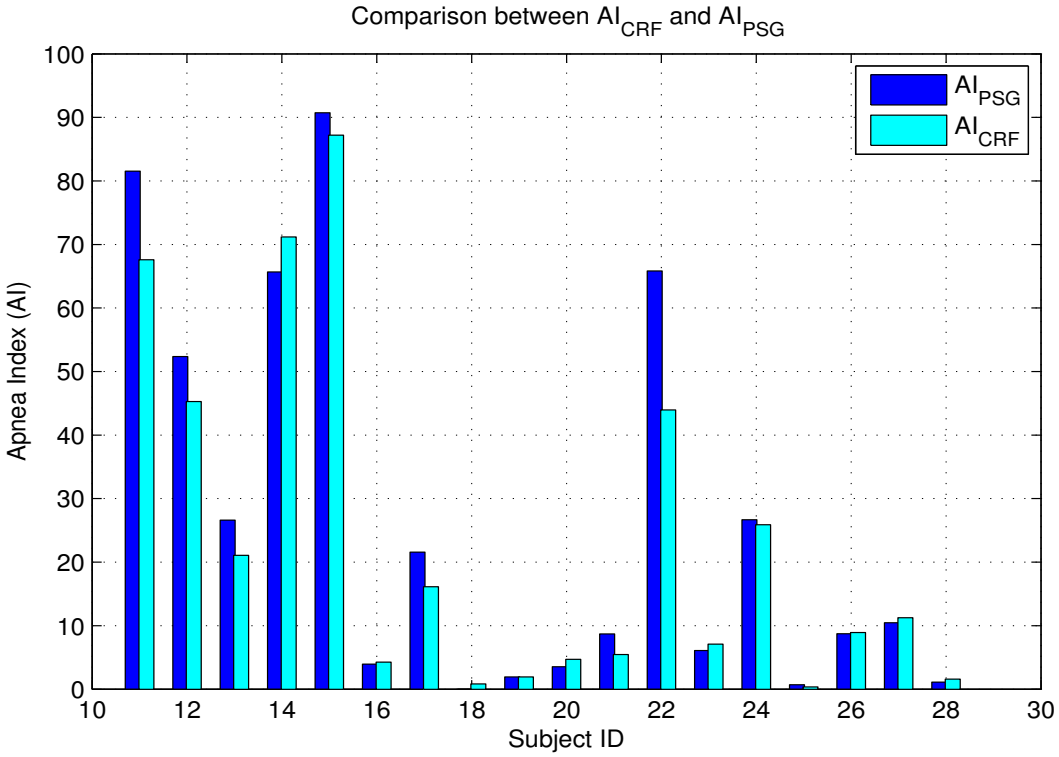


Figure 4.4: Comparison between AI_{CRF} and AI_{PSG}

shown in Table 4.5, the AIs calculated with detected apnea events can effectively categorize the 18 subjects with 92.31% sensitivity[2] and 80% specificity[3]. In this table, true positives (TP) and false negatives (FN) refer to patients correctly diagnosed as having OSA and incorrectly diagnosed as healthy, while true negatives (TN) and false positives (FP) refer to normal subjects correctly diagnosed as healthy and incorrectly diagnosed as having OSA, respectively.

Relevant information on the two misdiagnosed subjects (S19 and S26) is presented in Table 4.6. We find that the number of apnea events detected by CRF is indeed close approximations of that reported by PSG for both patients. However, for S19, misdiagnosis is the result of the large discrepancy in the number of hypopnea events. Despite

[2]sensitivity: $\frac{TruePositive}{TruePositive+FalseNegative}$, which represents the proportion of actual positives that are correctly identified as positives.

[3]specificity: $\frac{TrueNegative}{TrueNegative+FalsePositive}$, which represents the proportion of negatives that are correctly identified as negatives.

Subject	Method	AEN	HEN	AI	HI
S19	PSG	11	114	1.9	20
	CRF	11	1	1.9	0.1
S26	PSG	45	8	8.7	1.6
	CRF	46	1	8.9	0.2

Table 4.6: Information of incorrectly categorized subjects

effective apnea event detection, we are unable to identify OSA subjects whose dominant symptom is hypopnea due to our limited capacity in hypopnea event detection. S26 should have been correctly classified because the detected number for apnea events and hypopnea events are both close approximations. However, because the ground truth uses AHI = 15 as cutoff and CRF uses AI = 5, subjects with AHI < 15 and AI > 5, such as S26, would be incorrectly diagnosed. Nonetheless, our proposed method can still provide close approximations of the AI.

Actually, the total number of apnea events detected by CRF contains both the correctly detected ones and the incorrectly detected ones. Therefore, rough numbers listed in Table 4.4 are insufficient to illustrate the effectiveness of apnea event detection using the proposed relational learning approach. The number of correctly detected apnea events should be investigated.

Subject	AEN_{PSG}	AEN_{CRF}	correct AEN_{CRF}	Ratio of correctly AEN_{CRF}
S11	392	325	227	69.84%
S12	274	237	174	73.42%
S13	135	107	72	67.29%
S14	359	388	249	64.18%

Table 4.7: Comparison between AEN_{PSG} and correct AEN_{CRF}

Besides the 10 pieces of training data (S1–S10), four more pieces (S11–S14) have respiratory event annotations done by technicians. These annotations are used as the reference to measure the number of correctly detected apnea events using CRF. As shown

in Table 4.7, about 70% of the apnea events have been correctly detected. This is a promising result given that we are the first group that validates the performance on apnea event detection. Moreover, we have shown that automatic annotation of apnea events using CRF has the potential to replace manual annotation and reduce labor cost.

4.5 Comparison With Existing Diagnostic Methods

In this section, we compared the performance of CRF on OSA diagnosis with three existing methods, two of which are snore-episode-related and the other respiratory-event-related.

The two snore-episode-related methods are the ISPJ method and the F1 method. ISPJ and F1 have been claimed as effective indicators of OSA diagnosis, and our method adopts both of them as observations, but does it perform better than using either ISPJ or F1 individually?

The respiratory-event-related method is proposed by Hou and Xie et al. who claimed that respiratory events, i.e. apnea events and hypopnea events, can be detected by endpoint detection (EPD) method for snore episodes using only properties in the time domain[22]. Our method also incorporates features from the frequency domain and the sequential relation between long silence episodes and their neighboring snore episodes. Does our combined approach detect respiratory events better?

Therefore, we conducted comparison experiments to answer these questions.

4.5.1 Comparison With the Snore-Episode-Related Methods

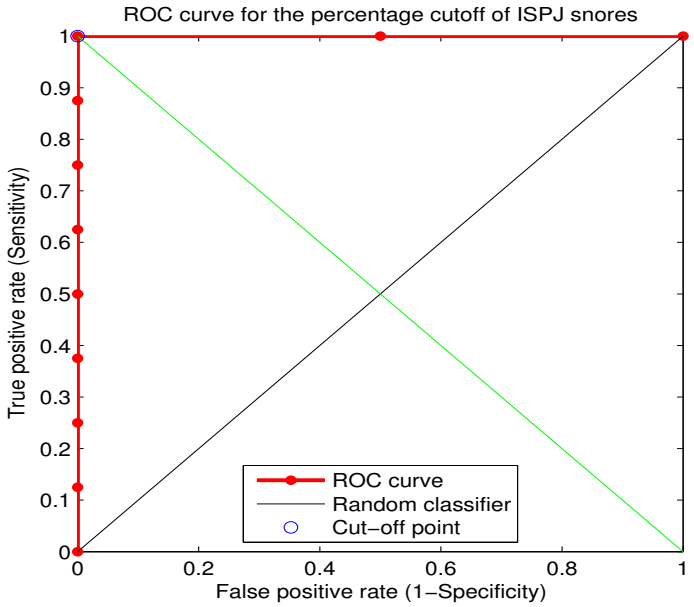


Figure 4.5: ROC curve for the percentage cutoff PP_{th} of snores labeled with ISPJ

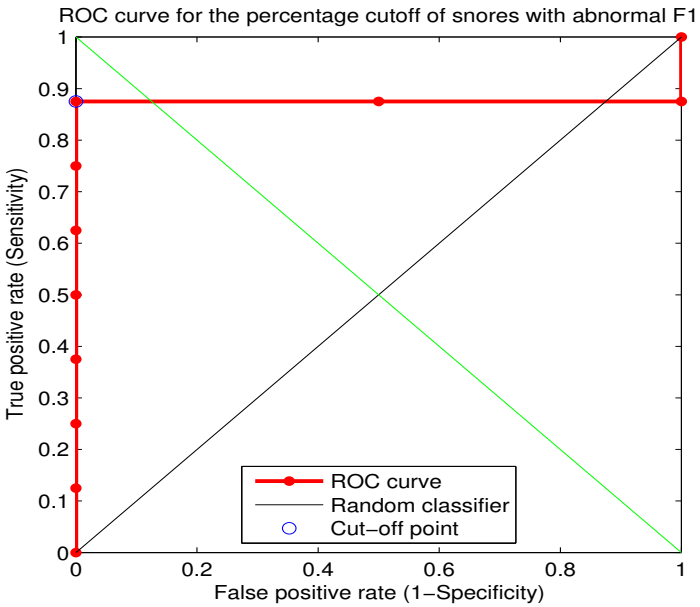


Figure 4.6: ROC curve for the percentage cutoff FP_{th} of snores labeled with abnormal F1

The ISPJ and F1 methods compute the percentage P_{ISPJ} and $P_{\text{F1}}$of snore episodes labeled with ISPJ or abnormal F1, respectively. We validated their diagnostic performance using our training and testing data and the parameters listed in Table 4.1. Let PP_{th} and FP_{th} be the percentage cutoffs for these two methods, respectively. An ROC analysis is used to derive PP_{th} and FP_{th} from the diagnosis results of the ten training pieces. Figure 4.5 shows the ROC curve for PP_{th}, which has the optimal decision threshold of 5.8% with 100% sensitivity and 100% specificity. Figure 4.6 shows the ROC curve for FP_{th} , which has the optimal decision threshold of 12.89% with 87.5% sensitivity and 100% specificity).

	ISPJ	F1	CRF
Optimal threshold	$PP_{th} = 5.8\%$	$FP_{th} = 12.89\%$	AI = 5
TP FP FN TN	11 3 2 2	11 2 2 3	12 1 1 4
Sensitivity	84.62%	84.62%	92.31%
Specificity	40%	60%	80%
Correlation with AHI_{PSG} (r)	0.830	0.737	0.879

Table 4.8: Performance of OSA diagnosis using ISPJ, F1, and CRF

The 18 pieces of testing data are then diagnosed with the derived percentage cutoffs, that is, a subject is diagnosed as having OSA if $P_{\text{ISPJ}} \geq PP_{th}$ for the ISPJ method or if $P_{\text{F1}} \geq FP_{th}$ for the F1 method. Table 4.8 shows the result of OSA diagnoses using ISPJ, F1, and CRF. As shown in this table, CRF classifies OSA patients and simple snorers more effectively than the other two methods. Our method, which combines the ISPJ and F1 labels, indeed performs better than using either ISPJ or F1 individually.

Ideally, the results of a diagnostic method should covary with the clinical measurement (AHI in the case of OSA) closely. Therefore, in Table 4.8, we also measured the correlation between AHI_{PSG} and P_{ISPJ}, P_{F1} and AI_{CRF}. As shown summarily in Table 4.8 and in detail in Figure 4.7, AI_{CRF} is best correlated with AHI_{PSG} (r_{CRF}=0.879 >

r_{ISPJ}=0.830 > r_{F1}=0.737). This finding illustrates that in addition to effectively providing binary classification, our method is also capable of a more detailed OSA diagnosis.

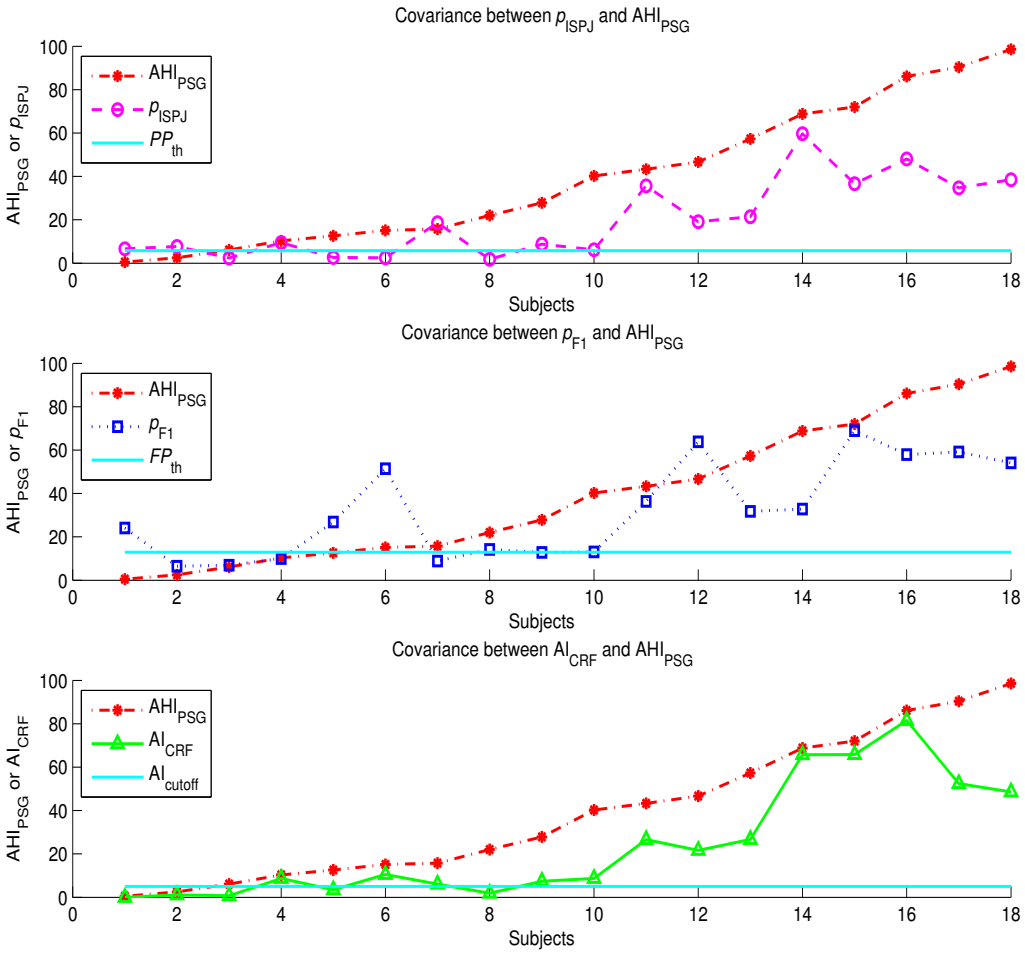


Figure 4.7: Covariance between AHI_{PSG} and P_{ISPJ}, P_{F1} and AI_{CRF}

4.5.2 Comparison With the Respiratory-Event-Related Method

To be consistent with previous comparison experiments, we only evaluated EPD on the 18 pieces of testing data (S11–S28). Hou and Xie et al. collectively identify apnea

events and hypopnea events as respiratory events in [22]. Therefore, we compared their result with the number of respiratory events detected by PSG in Figure 4.8. As shown in this figure, EPD detects considerably more respiratory events than PSG in 12 of the 18 cases, with an actual average difference of 85.75 events and a max difference even of 333 events. Such huge differences are the result of EPD's definition for respiratory events - silence episodes longer than 10 seconds between two sound episodes. This definition cannot eliminate any long periods of normal, silent breathing.

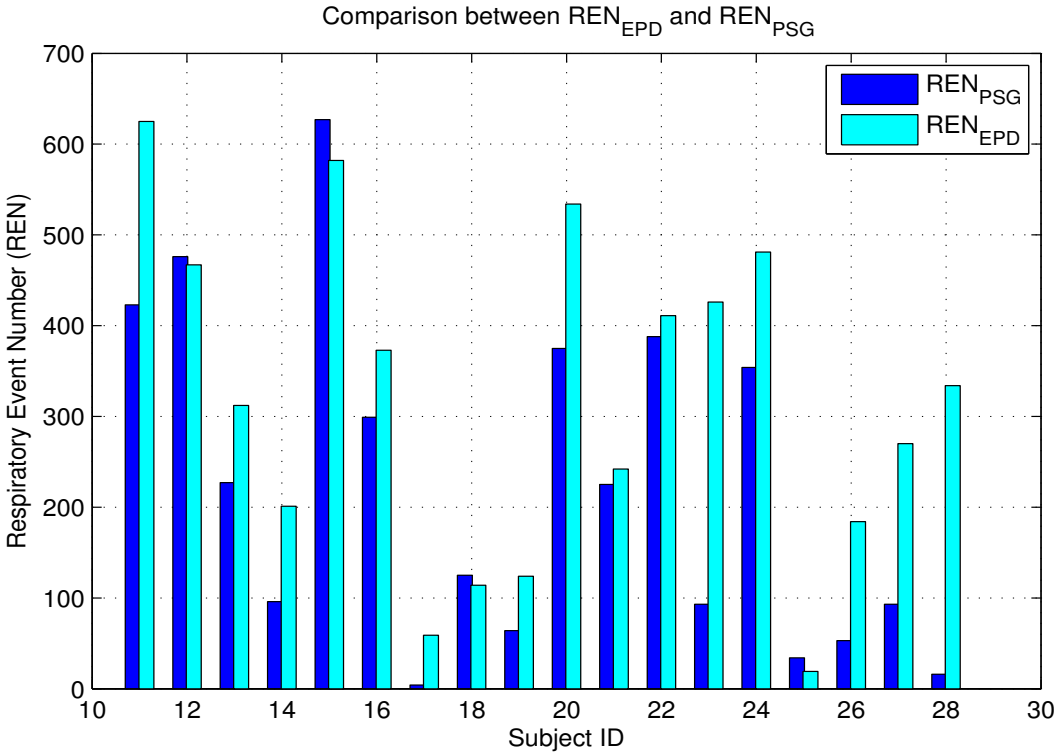


Figure 4.8: Comparison between REN detected by EPD and by PSG

Given that EPD uses the AHI value to diagnose OSA , AHI = 15 is adopted as its threshold in the comparison experiment. As shown in Table 4.9, CRF categorizes simpler snores more effectively than EPD, but it is less effective in identifying OSA patients. On the other hand, EPD can identify all of the OSA patients but often over-estimates the subjects' severity because it tends to detect more respiratory events than

	EPD	CRF
Threshold	AHI = 15	AI = 5
TP FP FN TN	13 3 0 2	12 1 1 4
Sensitivity	100%	92.31%
Specificity	40%	80%
Correlation with AHI_{PSG} (r)	0.834	0.879

Table 4.9: Comparison between OSA diagnosis results of EPD and CRF

the actual number (14 in 18 as shown in Figure 4.8). Our method, however, can better differentiate apnea silences versus normal silences by combining features from both the time and the frequency domains and the sequential relation among adjacent acoustic signal components.

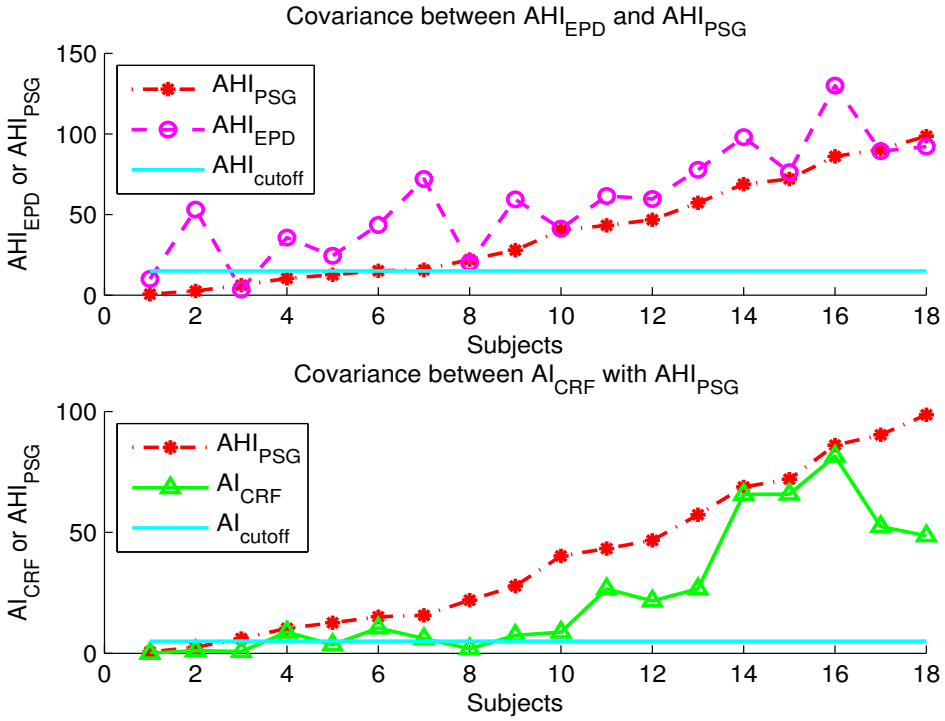


Figure 4.9: Covariance between AHI_{EPD} and AHI_{PSG} and that between AHI_{PSG} and AI_{CRF}

Table 4.9 also indicates that AI_{CRF} has a better correlation with AHI_{PSG}than AHI_{EPD} (r_{CRF}=0.879 > r_{EPD}=0.834). Figure 4.9 visualizes the covariance between AHI_{EPD} and AHI_{PSG}, as well as that between AI_{CRF} and AHI_{PSG}. Actually, AHI_{EPD} should be more correlated and closer to AHI_{PSG} since it is expressed in terms of AHI. However, as shown in Figure 4.9, large discrepancy exists between the curve of AHI_{EPD} and AHI_{PSG}, and it does not follow the general trend of $AHI_{PSG.}$ Moreover, Figure 4.9 clearly shows that EPD almost always detects a higher AHI than the actual one, which provides an intuitive explanation for why EPD can perfectly categorize OSA patients, but fails to identify most healthy subjects. Although the AI determined by our method tends to underestimate the severity of OSA, it is reasonable given that HI is left without consideration.

In this section, several experiments were conducted to learn the best parameter setting and train the CRF model. Moreover, we also validated the performance of our method in apnea events detection and OSA diagnosis by comparing with two snore-episode-related methods and one respiratory-event-related method. The results showed that our method exhibited the best performance by achieving a sensitivity of 92.31% and a specificity of 80%.

5 Mobile Obstructive Sleep Apnea Diagnosis

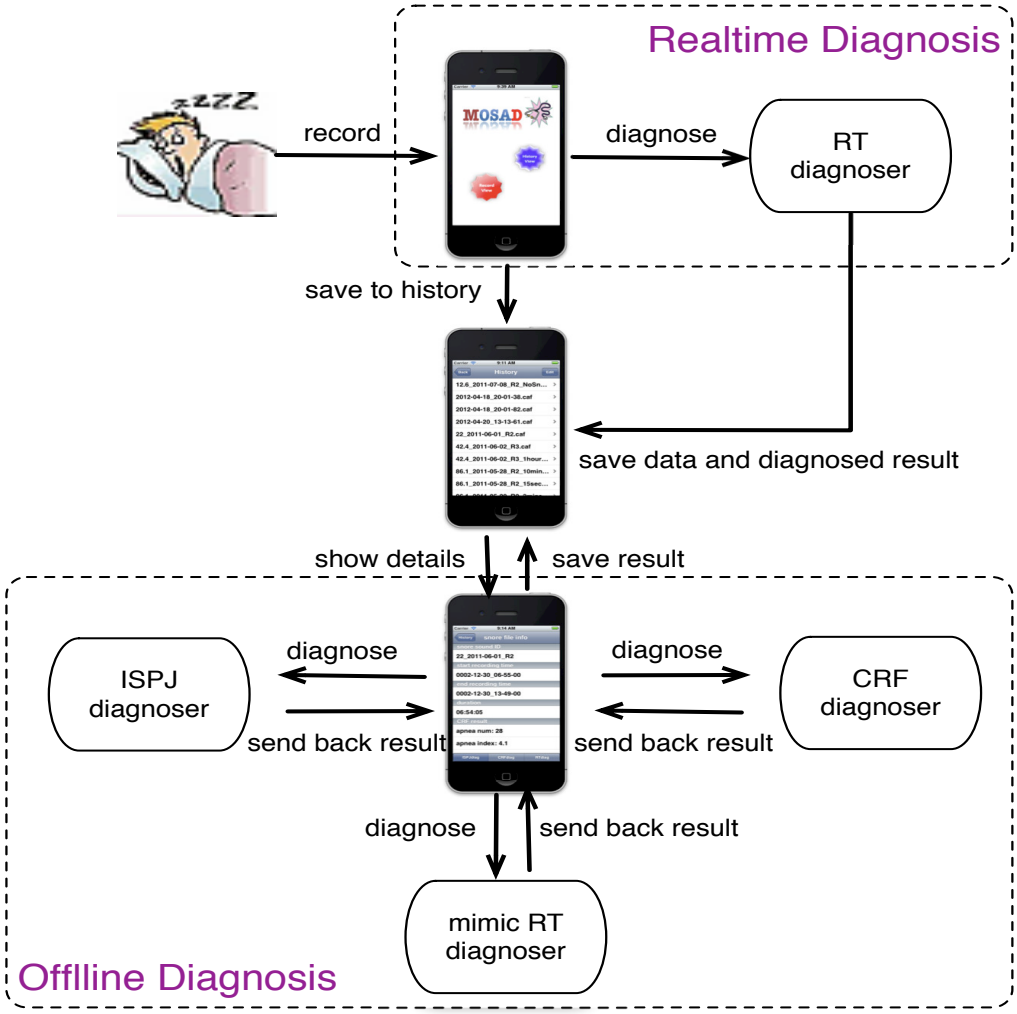


Figure 5.1: MOSAD prototype

We developed a MOSAD prototype shown in Figure 5.1 on iOS as a potential home screening tool for OSA. This prototype supports the following functionalities:

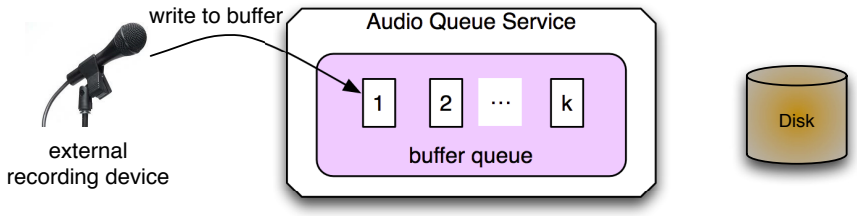
- Recording the overnight snore-related sound;
- Extracting diagnostic information in real-time;

- Presenting diagnostic result once the recording process finishes;
- Maintaining a list of past recorded signals and diagnostic results;
- Selecting diagnostic methods: ISPJ and CRF;
- Saving the result of different diagnostic methods;
- Stopping and re-starting diagnosis.

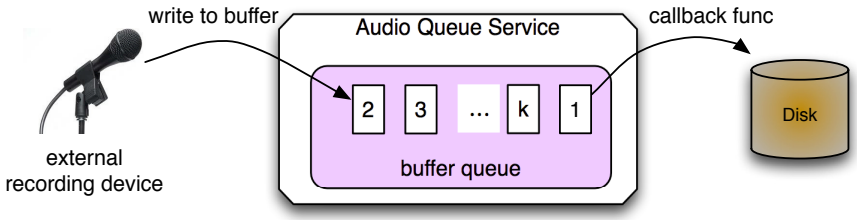
As shown in Figure 5.1, the prototype consists of two parts: real-time OSA diagnosis and offline OSA diagnosis. Both parts require the recording of snoring sound, and the optimization of audio processing are also applied in these two parts. Therefore, in the following sections, we will first describe the implementation of recording, followed by the optimization of audio processing. Then we will introduce the real-time and offline diagnosis, and finally we present result from the validation experiment on the mobile device.

5.1 Recording of Snore-related Signal

The Audio Queue Service in the Audio Toolbox framework offered by iOS is used to record the snoring sound. The core of Audio Queue Service is a queue with k ($k \geq 1$) buffers. The incoming audio from external audio hardware such as a microphone is stored in the first buffer at the head of the queue (Figure 5.2a). Once the first buffer fills up, it will be moved to the end of the queue, and a callback function will write the data stored in the buffer to the destination (Figure 5.2b). The data is thus flushed to the disk,



(a) Fill buffer with data from external recording device



(b) Write to disk via callback function

Figure 5.2: Recording audio queue

and the buffer is cleared for the next cycle of storage. Meanwhile, the incoming audio data continues to fill the subsequent buffers in the queue.

5.2 Optimization of Audio Processing

The two major audio processing algorithms used in our method are the YIN algorithm and the LPC-based formant calculation algorithm. When implemented on mobile devices, they create large delays due to limited CPU and memory capacity. It takes about 20 minutes to process a three-minute audio clip for offline diagnosis, and while audio is being processed, playback is not smooth. Real-time diagnosis will risk the loss of data once all the buffers in the queue are filled and cannot be released because of the time-consuming audio processing. To improve the efficiency of pitch determination and formant calculation, the algorithms of audio processing are optimized from two aspects: reducing time complexity of audio processing methods and avoiding redundant audio analyses.

5.2.1 Reduce Time Complexity of Audio Processing

YIN algorithm is a well-known method of pitch determination. It modifies auto-correlation to improve the accuracy of pitch calculation. However, its time complexity is still $O(n^2)$, which results in significant lag time. Because the basic theory of almost all methods in time domain is auto-correlation, all of these methods have the same time complexity as the YIN algorithm. The LPC-based formant calculation method also costs no less than $O(n^2)$. To speed up audio processing, more efficient techniques are investigated.

Cepstrum is a technique for signal processing [9]. The reason that cepstrum-based pitch determination and formant extraction method performs more efficiently is the application of the Fast Fourier Transform (FFT) technique, which can reduce the time cost to $O(n \log n)$. As shown in Figure 5.3, cepstrum first transfers the signals from the time domain to the frequency domain using the Fourier transform, generating the so-called spectrum. The spectrum is then converted to the cepstrum by using the inverse Fourier transform. The independent variable in cepstrum is called quefrency and the domain of cepstrum is then called the quefrency domain.

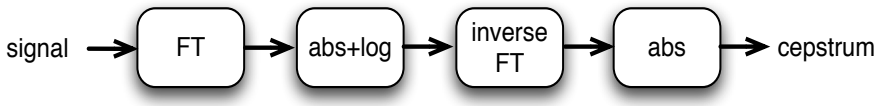


Figure 5.3: Cepstrum calculation

The motivation of these conversions is to separate features that depict the excitation source and the vocal tract in the quefrency domain. Previous research shows that the characteristic of the vocal excitation, pitch, and that of the vocal tract, formant frequencies, are embedded respectively in the low and high quefrency domain [24, 36]. Therefore, once the signal is transformed into quefrency domain, pitch and formant can

then be easily extracted.

As introduced in Section 3.3.4, snoring sound $f(n)$ can be decomposed into the convolution of vocal excitation $s(n)$ and vocal tract $h(n)$. When the signal is transferred into the frequency domain, the convolution is converted into multiplication:

$$F(\omega) = S(\omega) \cdot H(\omega),$$

$F(\omega)$, $S(\omega)$ and $H(\omega)$ are the corresponding Fourier Transform of $f(n)$, $s(n)$, and $h(n)$. The following equation computes the magnitude of the spectrum:

$$|F(\omega)| = |S(\omega)| \cdot |H(\omega)|.$$

After taking the log operation, the multiplication of spectrum components of the excitation source and the vocal tract is transformed into a linear combination of these components:

$$\log|F(\omega)| = \log|S(\omega)| + \log|H(\omega)|.$$

The separation of these two components is then done by taking the inverse Fourier transform of the log spectrum of the original signal.

The lifter operation in the cepstrum domain is similar to the filter in the frequency domain. A low time lifter intercepts the cepstrum component in the low quefrency region while the high time lifter only keeps those in the high quefrency region. Therefore, a lifter operation can separate cepstrum for pitch determination and formant extraction. Then, pitch can be obtained by using a peak picking algorithm in the high quefrency region. To extract formant, another Fourier Transform is conducted on cepstrum components in the low quefrency region to generate the spectrum without complication from high frequency components, and smoothing techniques are then applied to the newly

generated spectrum. Afterwards, the local maxima are extracted sequentially as formants.

Figure 5.4 presents the steps of pitch determination and formant extraction using cepstrum.

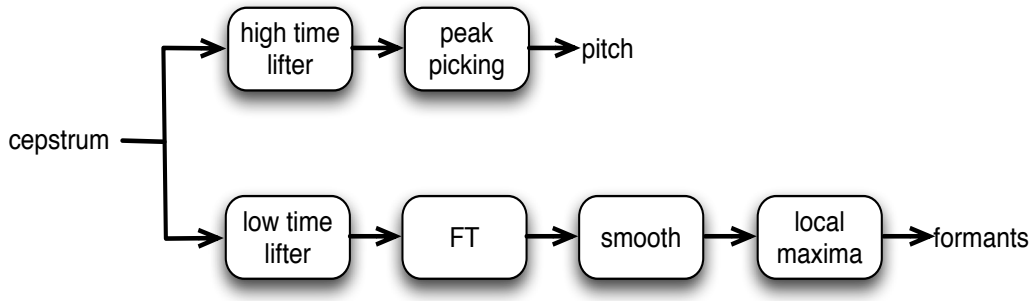


Figure 5.4: Extraction of pitch and formants from cepstrum

5.2.2 Avoid Redundant Audio Analysis

Although the efficiency of audio processing is enhanced dramatically using FFT, we can go even further. Based on the description of the ISPJ and F1 labels, a snore episode is tagged with the ISPJ and F1 label if there is at least one frame that has a pitch period and an F1 value above thresholds. Therefore, the calculation of the pitch and the formant is not necessary for every frame in a snore episode. Once one frame is determined with ISPJ or abnormal F1, the whole episode can be tagged as such. When the whole episode has both labels, no further audio processing is required for the remaining frames. Such redundancy reduction is instrumental in creating efficient real-time diagnosis, as incoming buffers no longer require processing if they follow an episode already labeled with ISPJ and F1.

5.3 Estimation of Total Sleeping Time

TST refers to the total time from falling asleep to waking up, not the total time in bed. In PSG, TST can be easily detected with sensors to monitor the brain activity. However, it is difficult to detect TST merely from snoring sound. Thus, we approximate the TST with statistics shown in Table 5.1, which summarizes the averages and standard deviations (StdDev) of TST detected by PSG (TST_{PSG}) and the duration of recording (RecDur) for the 28 pieces of data.

	Mean (mins)	StdDev (mins)
TST_{PSG}	330.6	48.1
RecDur	432.3	19.7

Table 5.1: Statistics to estimate total sleeping time

We define the estimated TST (eTST) for a recording with the duration of *dur* as follows:

$$eTST = Mean(TST_{PSG}) + \frac{dur - Mean(RecDur)}{StdDev(RecDur)} \times StdDev(TST_{PSG}).$$

This formula is used to estimate the TST of each subject.

5.4 Real-time Diagnosis

Real-time diagnosis is conducted simultaneously with recording. With the audio queue structure introduced above, buffers in the queue continuously shuttle audio data from the external recording hardware to the disk. Meanwhile, the callback function processes the audio data in each buffer. The procedure of feature extraction is embedded in this callback function to extract diagnostic information.

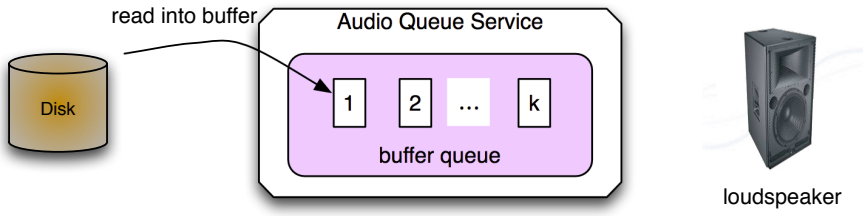
The data in each buffer is first tagged with a label (silence, snore, or non-snore), which is then used to decide whether to add the data as a new episode or attach it to the current episode. If a snore frame is added as a new episode, its pitch and F1 value are then calculated to generate the ISPJ and F1 labels, respectively. In contrast, if a snore frame is attached to the previous episode, whether to calculate its pitch and its F1 value depends on whether the previous episode has an ISPJ label and an F1 label. The observations for CRF will have already been obtained once the recording is finished. The CRF model is then applied to detect apnea events.

Ideally, the MOSAD prototype should be placed inside the sleeping laboratory's environment and made to perform the real-time diagnosis during actual PSGs for performance evaluation. However, recollecting data takes a long time. Since the data we used in previous experiments are also collected using the recording part in Section 5.1, we simulate the procedure of real-time diagnosis by using the playback Audio Queue Service on existing data. The structure of the playback audio queue is quite similar to that of the recordings.

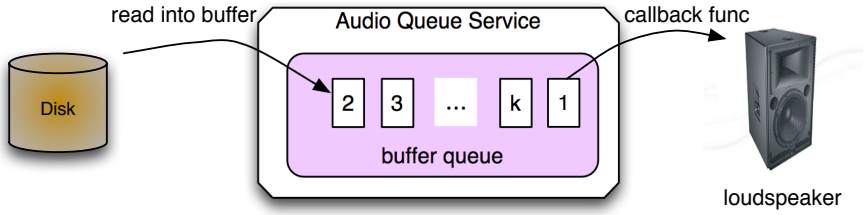
As shown in Figure 5.5, the difference between the recording and the playback audio queue structure lies in the incoming source and the outgoing destination. The former shuttles audio data from external recording devices to disk while the latter shuttles from disk to external playback devices. However, both of them store data in a buffer queue in memory and make use of the callback function once the buffer is filled up. Thus, we simulate real-time diagnosis by utilizing the playback audio queue to read the existing signal.

5.5 Offline Diagnosis

After the data is collected and stored, users can choose amongst three diagnostic methods: ISPJ, offline CRF, and "real-time" CRF.



(a) Fill buffer by reading data from disk



(b) Playback via callback function

Figure 5.5: Playback audio queue

The ISPJ method calculates the percentage of snore episodes labeled with ISPJ.

Although there are two CRF methods, offline diagnosis accesses snoring sound in a different manner. It adopts the Extended Audio File Service to sequentially read a chunk, which is much larger than the size of an Audio Queue buffer, from the recorded audio each time. Each chunk is first segmented into non-snore, snore, and silence episodes, from which diagnostic information are then extracted. When the entire piece of data is processed, a merge step is taken to link the boundaries between chunks, and the diagnostic information is then fed to the CRF to identify the apnea events.

5.6 Validation Experiments

5.6.1 Specification of Mobile Device

The validation experiment uses an iPod Touch 4G with the following specifications:

iPod Touch 4g	
installed OS	iOS 5.0
CPU	ARM Cortex-A8 Apple A4 1 GHz (underclocked to 800 MHz)
Storage	32 GB flash memory
Memory	256 MB DRAM

Table 5.2: Specifications of the iPod Touch

.

As shown in Table 5.2, the iPod Touch's memory capacity and CPU processing speed are the major limitations for OSA diagnosis and should thus be considered when designing mobile applications.

5.6.2 Performance Validation on Mobile Device

To evaluate the performance of the MOSAD prototype on mobile devices, all of the 18 pieces of recording are tested on the iPod Touch using real-time CRF diagnosis and offline CRF diagnosis, both of which perform consistently in classification. Table 5.3 shows the performance of CRF implemented on PC (CRF_{PC}) and the one implemented on iOS (CRF_{iOS}).

	CRF_{iOS} (real-time diagnosis and offline diagnosis)	CRF_{PC}
Optimal threshold	AI = 5	AI = 5
TP FP FN TN	9 0 4 5	12 1 1 4
Sensitivity	69.23%	92.31%
Specificity	100%	80%

Table 5.3: Performance of OSA diagnosis using CRF on iOS

Compared to CRF_{PC}, CRF_{iOS} performs better in identifying healthy subjects, but is less effective in identifying OSA patients. Given the limitations of the iPod Touch's

computing power, this discrepancy may arise from the following adaptations:

- Different thresholds for segmentation. In CRF_{iOS}, the signal is recorded in linear PCM, which represents each sample with a 16-bit integer. In contrast, the signal in CRF_{PC} is transformed into a wave file, and each sample is a float number between 1 and -1. The inconsistent representation of samples causes the derivation of different thresholds, consequently affecting the performance of segmentation.

- Different chunk sizes. incoming signals are read and processed in 25M (five-minute) chunks in CRF_{PC} but 10M chunks in CRF_{iOS} due to the limited memory capacity of mobile devices. The chunk size will affect the segmentation because small chunks introduce more fragments.

- Different pitch extraction and F1 determination methods. CRF_{PC} utilizes the YIN algorithm and the LPC-based method to determine pitch and formant, respectively, but CRF_{iOS} replaces these two methods with cepstrum calculation for efficiency. These two methods may yield different pitch and formant values, which affects the training of CRF model and the inference of CRF in apnea event detection.

- Different TST. CRF_{PC} adopts the actual TST_{PSG} while CRF_{iOS} estimates TST based on the duration of recording and the statistics collected from existing recordings. The gap between the TST_{PSG} and the eTST affects the calculation of AHI.

5.6.3 Efficiency Experiments

In addition to performance, efficiency is a major factor in the mobile application design. The most time-consuming part of the prototype is the calculation of pitch and

formant. Using the original pitch determination and formant extraction method, offline diagnosis on one piece of recording can be completed in approximately one hour on the simulator on PC, but when running on an actual mobile device, it costs considerably more time. As discussed before, even a three-minute clip requires almost 20 minutes to process. This renders offline diagnosis using complete overnight recordings unfeasible. However, after the optimization of audio processing, the average time for offline CRF diagnosis is 88 ± 44 minutes, which is completely feasible and far more reasonable considering the long duration of a whole-night recording.

For real-time CRF diagnosis, the bottleneck is not time but power usage. A fully charged iPod Touch can power three to four overnight recordings, but does not last even one night with real-time diagnosis. Thus, the device must stay plugged in during the process. The large amount of calculation for pitch and formant accounts for such power consumption. Methods with less calculation should be investigated in the future.

6 Conclusion and Future Work

In this thesis, we proposed a relational learning method to detect apnea events using CRF, a well-known sequential labeling technique. The proposed method combines features from the time domain, which capture the typical event pattern for apnea, and features from the frequency domain, which carry useful information about snore source and upper airway abnormalities. To identify respiratory events, CRF determines the relations of these features among adjacent acoustic signal components, including silence episodes and snore episodes. Experiments demonstrated the effectiveness of our method in apnea event detection. Furthermore, compared with two existing snore-episode-related and one respiratory-events-related diagnostic methods, CRF performed better in the classification of OSA patients and healthy subjects. Our proposed approach thus has the potential to relieve sleep technicians of the burden of manual annotation.

We also developed MOSAD, a prototype of OSA diagnosis on iOS, which can record snoring sound, provide diagnoses in both real-time and offline mode, and enable users to view past records and diagnosis results. Validation experiments showed that both real-time and offline diagnosis could effectively identify healthy subjects but had room for improvement when it came to identifying OSA patients. Experiments also demonstrated the reasonable efficiency of offline diagnosis. Due to limited time and human resources, validation experiments were not conducted in the home environment. However, the performance based on the recording collected from the sleeping laboratory shows MOSAD's promise as an effective home screening tool for OSA.

Although the proposed method has been shown to effectively diagnose OSA, and preliminary validation demonstrates the potential of MOSAD prototype as a OSA home screening tool, we can still make our method and our prototype more effective and applicable by:

- **Improving the accuracy of apnea detection**. Currently, CRF only uses four observations for the training and inference, and it only learns the relation between snore episodes and silences episodes. Additional features and relation among other acoustic signal components, such as breath episodes, can be investigated.

- **Investigating features to identify hypopnea events**. The CRF method cannot effectively detect hypopnea events because no obvious pattern for hypopnea has been observed in snoring signals. Further research on finding typical patterns and features of hypopnea events in snoring sound should be carried out in the future.

- **Validating the performance in home environment**. Although the MOSAD prototype is validated with data recorded in the ideal environment of sleeping laboratory, home environment may contain various background noise. The squeaking of the bed, noise from the air conditioners, the clocks, rotating fans and sounds produced by the bed partner may interfere with the OSA diagnosis.

- **Applying noise reduction techniques**. The data we used was recorded in the sleeping laboratory, which was with noise level intentionally reduced. In actual home environments, noise reduction techniques such as time-frequency filter and source separation should be applied to pre-process the signal before further analyses.

References

[1] U. R. Abeyratne, C. K. K. Patabandi, and K. Puvanendran. Pitch-jitter analysis of snoring sounds for the diagnosis of sleep apnea. 2:2072–2075 vol.2, 2001.

[2] Udantha Abeyratne, Asela Karunajeewa, and Craig Hukins. Mixed-phase modeling in snore sound analysis. *Medical and Biological Engineering and Computing*, 45(8):791–806, August 2007.

[3] Udantha R. Abeyratne, Ajith S. Wakwella, and Craig Hukins. Pitch jump probability measures for the analysis of snoring sounds in apnea. *Physiological Measurement*, 26(5):779–798, October 2005.

[4] D. Alvarez, R. Hornero, M. García, F. del Campo, and C. Zamarrón. Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure. *Artificial intelligence in medicine*, 41(1):13, 2007.

[5] I. Ayappa and D.M. Rapoport. The upper airway in sleep: physiology of the pharynx. *Sleep Medicine Reviews*, 7(1):9–33, 2003.

[6] R. Beck, M. Odeh, A. Oliven, and N. Gavriely. The acoustic properties of snores. *European Respiratory Journal*, 8(12):2120–2128, 1995.

[7] Cavusoglu, M., Ciloglu, T., Serinagaoglu, Y., Kamasak, M., Erogul, O., Akcam, and T. Investigation of sequential properties of snoring episodes for obstructive sleep apnoea identification. *Physiological Measurement*, 29(8):879–898, August 2008.

[8] Cavusoglu, M., Kamasak, M., Erogul, O., Ciloglu, T., Serinagaoglu, Y., Akcam, and T. An efficient method for snore/nonsnore classification of sleep sounds. *Physiological Measurement*, 28(8):841–853, August 2007.

[9] D.G. Childers, D.P. Skinner, and R.C. Kemerait. The cepstrum: A guide to processing. *Proceedings of the IEEE*, 65(10):1428–1443, 1977.

[10] F. Chung, B. Yegneswaran, P. Liao, S.A. Chung, S. Vairavanathan, S. Islam, A. Khajehdehi, and C.M. Shapiro. Stop questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology*, 108(5):812–821, 2008.

[11] F. Chung, B. Yegneswaran, P. Liao, S.A. Chung, S. Vairavanathan, S. Islam, A. Khajehdehi, and C.M. Shapiro. Validation of the berlin questionnaire and american society of anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients. *Anesthesiology*, 108(5):822–830, 2008.

[12] A. De Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am*, 111(4):1917–1930, 2002.

[13] D. DeCaprio, J.P. Vinson, M.D. Pearson, P. Montgomery, M. Doherty, and J.E. Galagan. Conrad: gene prediction using conditional random fields. *Genome research*, 17(9):1389–1398, 2007.

[14] Charalampos Doukas, Theodoros Petsatodis, and Ilias Maglogiannis. MORFEAS: A Non-Invasive System for automated Sleep Apnea Detection utilizing Snore Sound Analysis. In Ratko Magjarevic, Panagiotis D. Bamidis, and Nicolas Pallikarakis, editors, *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, volume 29 of *IFMBE Proceedings*, chapter 30, pages 119–123. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[15] Duckitt, W.D., Tuomi, S.K., Niesler, and T.R. Automatic detection, segmentation and assessment of snoring from ambient acoustic data. *Physiological Measurement*, 27(10):1047–1056, October 2006.

[16] Takahiro Emoto, Udantha R. Abeyratne, Masatake Akutagawa, Asela S. Karunajeewa, Shinsuke Konaka, and Yohsuke Kinouchi. Comparing benign snores with apneic snores based on high frequency analysis. In Ratko Magjarevic, Olaf Dössel, and Wolfgang C. Schlegel, editors, *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, volume 25/7 of *IFMBE Proceedings*, chapter 227, pages 820–823. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[17] R. Ferber, R. Millman, M. Coppola, J. Fleetham, C.F. Murray, C. Iber, V. McCall, G. Nino-Murcia, M. Pressman, M. Sanders, et al. Portable recording in the assessment of obstructive sleep apnea. asda standards of practice. *Sleep*, 17(4):378, 1994.

[18] W.W. Flemons, M.R. Littner, J.A. Rowley, P. Gay, W.M.D. Anderson, D.W. Hudgel, R.D. McEvoy, and D.I. Loube. Home diagnosis of sleep apnea: A systematic review of the literature*. *Chest*, 124(4):1543–1579, 2003.

[19] GJ Gibson. Obstructive sleep apnoea syndrome: underestimated and undertreated. *British medical bulletin*, 72(1):49–64, 2004.

[20] X. He, R.S. Zemel, and M.A. Carreira-Perpinán. Multiscale conditional random fields for image labeling. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–695. IEEE, 2004.

[21] Michael Herzog, Andreas Schmidt, Thomas Bremert, Beatrice Herzog, Werner Hosemann, and Holger Kaftan. Analysed snoring sounds correlate to obstructive sleep disordered breathing. *European archives of oto-rhino-laryngology*, 265(1):105–113, January 2008.

[22] L. Hou, S. Xie, S. Kai, and W. Song. Detection of osahs using only time-domain property of snoring signal. In *Multimedia Technology (ICMT), 2011 International Conference on*, pages 519–522. IEEE, 2011.

[23] R. Jane, J. A. Fiza, J. Sola-Soler, S. Blanch, P. Artis, and J. Morera. Automatic snoring signal analysis in sleep studies. pages 366–369, 2003.

[24] M.A. Kammoun, D. Gargouri, M. Frikha, and A. Ben Hamida. Cepstral method evaluation in speech formant frequencies estimation. In *Industrial Technology, 2004. IEEE ICIT'04. 2004 IEEE International Conference on*, volume 3, pages 1612–1616. IEEE, 2004.

[25] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins. Pitch/Airway-Response and Cepstral Analysis of Snore Sounds for the Non-Contact Screening of Sleep Apnea. 25/4:2295–2298, 2010.

[26] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins. Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome. *Physiological measurement*, 32:83, 2011.

[27] S.M. Khoo, H.K. Poh, Y.H. Chan, W.J. Ngerng, D.X. Shi, and TK Lim. Diagnostic characteristics of clinical prediction models for obstructive sleep apnea in different clinic populations. *Sleep and Breathing*, 15(3):431–437, 2011.

[28] R. Klinger and K. Tomanek. *Classical probabilistic models and conditional random fields*. TU, Algorithm Engineering, 2007.

[29] J. Lafferty, A. McCallum, and F.C.N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[30] Z. Lan, A. Itoi, M. Takashima, M. Oda, and K. Tomoda. Difference of pharyngeal morphology and mechanical property between osahs patients and normal subjects. *Auris Nasus Larynx*, 33(4):433–439, 2006.

[31] N.C. Netzer, R.A. Stoohs, C.M. Netzer, K. Clark, K.P. Strohl, et al. Using the berlin questionnaire to identify patients at risk for the sleep apnea syndrome. *Annals of internal medicine*, 131(7):485, 1999.

[32] A. K. Ng, T. S. Koh, U. R. Abeyratne, and K. Puvanendran. Investigation of obstructive sleep apnea using nonlinear mode interactions in nonstationary snore signals. *Annals of biomedical engineering*, 37(9):1796–1806, 2009.

[33] Andrew K. Ng, T. S. Koh, Eugene Baey, and K. Puvanendran. Diagnosis of Obstructive Sleep Apnea using Formant Features of Snore Signals. In Ratko Magjarevic, R. Magjarevic, and J. H. Nagel, editors, *World Congress on Medical Physics and Biomedical Engineering 2006*, volume 14 of *IFMBE Proceedings*, chapter 230, pages 967–970. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[34] Andrew K. Ng, Tong S. Koh, Eugene Baey, Teck H. Lee, Udantha R. Abeyratne, and Kathiravelu Puvanendran. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine*, 9(8):894–898, December 2008.

[35] Ashida Nobuyuki, Nasu Yasuhiro, Teshima Taiki, Yamakawa Miyae, Makimoto Kiyoko, and Higashi Terumasa. Trial of measurement of sleep apnea syndrome with sound monitoring and SpO2 at home. pages 66–69, December 2009.

[36] A.M. Noll. Cepstrum pitch determination. *The journal of the acoustical society of America*, 41:293, 1967.

[37] Carole Planès, Michel Leroy, Bouach, El, Franck Digne, Florence de Roquefeuil, and Bernadette Raffestin. Home diagnosis of obstructive sleep apnoea in coronary patients: validity of a simplified device automated analysis. *Sleep and Breathing*.

[38] K. Puvanendran and K. L. Goh. From snoring to sleep apnea in a Singapore population. *Sleep research online : SRO*, 2(1):11–14, 1999.

[39] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[40] F. Roche, J.M. Gaspoz, P. Minini, V. Pichot, D. Duverney, F. Costes, J.R. Lacour, J.C. Barthélémy, et al. Screening of obstructive sleep apnea syndrome by heart rate variability analysis. *Circulation*, 100(13):1411–1415, 1999.

[41] J.I. Salisbury and Y. Sun. Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique. *Medical engineering & physics*, 29(3):336–343, 2007.

[42] D. Schlosshan and M. Elliott. Sleep* 3: Clinical presentation and diagnosis of the obstructive sleep apnoea hypopnoea syndrome. *Thorax*, 59(4):347, 2004.

[43] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Towards automatic pitch detection in snoring signals. 4:2974–2976 vol.4, 2000.

[44] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Snoring sound intensity study with ambient and tracheal microphones. 2:2032–2035 vol.2, 2001.

[45] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Pitch analysis in snoring signals from simple snorers and patients with obstructive sleep apnea. 2:1527–1528 vol.2, 2002.

[46] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Spectral envelope analysis in snoring signals from simple snorers and patients with Obstructive Sleep Apnea. pages 2527–2530, 2003.

[47] J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera. Variability of snore parameters in time and frequency domains in snoring subjects with and without Obstructive Sleep Apnea. pages 2583–2586, 2005.

[48] Jordi Sola-Soler, Raimon Jane, Jose A. Fiz, and Jose Morera. Automatic classification of subjects with and without Sleep Apnea through snoring analysis. pages 6093–6096, August 2007.

[49] Z.F. Udwadia, A.V. Doshi, S.G. Lonkar, and C.I. Singh. Prevalence of sleep-disordered breathing and sleep apnea in middle-aged urban indian men. *American journal of respiratory and critical care medicine*, 169(2):168, 2004.

[50] S.B. Wang, A. Quattoni, L.P. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1521–1527. Ieee, 2006.

[51] P.R. Westbrook, D.J. Levendowski, M. Cvetinovic, T. Zavora, V. Velimirovic, D. Henninger, and D. Nicholson. Description and validation of the apnea risk evaluation system*. *Chest*, 128(4):2166–2175, 2005.

[52] A. Yadollahi and Z. Moussavi. Acoustic obstructive sleep apnea detection. pages 7110–7113, September 2009.

[53] M. Yin, S. Miyazaki, Y. Itasaka, Y. Shibata, T. Abe, A. Miyoshi, K. Ishikawa, and K. Togawa. A preliminary study on application of portable monitoring for diagnosis of obstructive sleep apnea. *Auris Nasus Larynx*, 32(2):151–156, 2005.

[54] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr. The occurrence of sleep-disordered breathing among middle-aged adults. *The New England journal of medicine*, 328(17):1230–1235, April 1993.

A Appendix

Group	Subject	AHI	AEN	HEN	AI	HI	TST (mins)	longest AE (mins)	longest HE (mins)
training	S1	33.6	100	107	16.2	17.4	369.5	61.7	64
	S2	75	356	49	65.9	9.1	324	49.8	40.3
	S3	69.6	263	153	43.9	25.7	355	78.6	96.8
	S4	94.7	44	392	9.4	85.3	275	48.1	85.3
	S5	40.8	175	107	25.3	15.5	413	60.6	90.4
	S6	18.4	3	62	0.9	17.6	211.5	22.5	60.3
	S7	14.5	16	61	3	11.5	298.5	102.4	81.1
	S8	55.9	44	272	8.1	47.8	326.5	35.4	48.7
	S9	1.9	1	11	0.2	1.7	386.5	16.9	41.3
	S10	24.8	88	47	16.2	8.6	314.5	83.6	72.2
testing	S11	86.1	392	23	81.5	4.6	288.5	69.7	60
	S12	90.4	274	202	52.4	38	314	52.6	56.3
	S13	43.3	138	89	26.6	16.7	304.5	43.6	45.6
	S14	68.8	359	16	65.7	2.9	327	63.9	40.3
	S15	98.6	307	320	48.6	50	379	85.6	39.7
	S16	46.7	137	162	21.6	25.1	375.5	73.4	48
	S17	0.5	0	4	0	0.5	353.5	-	41.1
	S18	22	11	114	1.9	20	338.5	23.9	80.3
	S19	12.6	18	46	3.5	9	305	43	66.1
	S20	27.8	25	71	7.4	20.4	203	39.8	60.3
	S21	40.2	51	174	8.7	28.8	352	49.6	90.8
	S22	72	355	33	65.8	6.1	323.5	96.8	67.3
	S23	15.7	36	57	6.1	9.6	355	30.2	42.6
	S24	57.3	165	189	26.7	30.6	371	59	76.2
	S25	6.1	4	30	0.7	5.4	333.5	43.3	51.2
	S26	10.3	45	8	8.7	1.6	309.5	68.8	47.5
	S27	15.1	65	28	10.5	4.5	373	46.4	37.4
	S28	2.5	7	9	1.1	1.4	377.5	51.3	79.7

Table A.1: Information about subjects used in experiments