DETECTION OF POST APNEA SOUNDS AND APNEA PERIODS FROM
SLEEP SOUNDS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ERSİN KARCI

IN PARTIAL FULFILLMENT FOR THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JANUARY 2011

Approval of the thesis:

DETECTION OF POST APNEA SOUNDS AND APNEA PERIODS FROM SLEEP SOUNDS

submitted by **ERSİN KARCI** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**   ─────────

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering**   ─────────

Assist. Prof. Yeşim Serinağaoğlu Doğrusöz
Advisor, **Electrical and Electronics Eng Dept., METU**   ─────────

Assoc. Prof. Tolga Çiloğlu
Co-advisor, **Electrical and Electronics Eng Dept., METU**   ─────────

Examining Committee Members:

Prof. Dr. B. Murat Eyüboğlu
Electrical and Electronics Engineering Dept., METU   ─────────

Assist. Prof. Yeşim Serinağaoğlu Doğrusöz
Electrical and Electronics Engineering Dept., METU   ─────────

Prof. Dr. Nevzat G. Gençer
Electrical and Electronics Engineering Dept., METU   ─────────

Prof. Dr. Osman Eroğul
Biomedical Engineering Dept., GMMA   ─────────

Assoc. Prof. Tolga Çiloğlu
Electrical and Electronics Engineering Dept., METU   ─────────

Date: 27.01.2011

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Ersin KARCI

Signature:

ABSTRACT

DETECTION OF POST APNEA SOUNDS AND APNEA PERIODS FROM SLEEP SOUNDS

Karcı, Ersin

M.Sc., Department of Electrical and Electronics Engineering

Supervisor: Assist. Prof. Yeşim Serinağaoğlu Doğrusöz

Co-supervisor: Assoc. Prof. Tolga Çiloğlu

January 2011, 92 Pages

Obstructive Sleep Apnea Syndrome (OSAS) is defined as a sleep related breathing disorder that causes the body to stop breathing for about 10 seconds and mostly ends with a loud sound due to the opening of the airway. OSAS is traditionally diagnosed using polysomnography, which requires a whole night stay at the sleep laboratory of a hospital, with multiple electrodes attached to the patient's body. Snoring is a symptom which may indicate presence of OSAS; thus investigation of snoring sounds, which can be recorded in the patient's own sleeping environment, has become popular in recent years to diagnose OSAS. In this study, we aim to develop a new method to detect post-apnea snoring episodes with the goal of diagnosing apnea or creating a new criteria similar to apnea / hypopnea index. In this method, first segmentation is done to eliminate the silence parts and only deal with active. Then these episodes are represented by distinctive features; some of these features are available in literature but some of them are novel. Finally, these episodes are classified using supervised and unsupervised methods. We are especially interested in detecting post apnea episodes, hence the apnea periods. False alarm rates are reduced by adding additional constraints into the detection algorithm. These methods are applied to snoring sound signals of OSAS patients,

recorded in Gulhane Military Medical Academy, to verify the success of our algorithms.

ÖZ

UYKU SESLERİNDEN APNE SONRASI SESLERİN VE APNE ZAMANLARININ
BULUNMASI

Karcı, Ersin

Yüksek Lisans, Elektrik Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Yard. Doç. Dr. Yeşim Serinağaoğlu Doğrusöz

Ortak Tez Yöneticisi: Doç. Dr. Tolga Çiloğlu

Ocak 2011, 92 Sayfa

Tıkayıcı uyku apnesi sendromu (TUAS), vücutta nefes alıp verememenin yaklaşık 10
saniye durmasına sebep olan uyku ile ilişkili bir solunum sorunudur. TUAS tanısı
standart olarak hastanın bütün gece uyku laboratuarında bir çok elektroda bağlı
olarak kalmasını zorunlu kılan polisomnografi cihazı ile teşhis edilir. Horlama, TUAS
için bir belirti olabileceğinden, son yıllarda hastanın kendi uyku ortamında kolayca
kaydedilebilen horlama sesleri, TUASın tanısının konmasında popüler bir hale geldi.
Biz bu çalışmada horlama seslerinden apne zamanlarını yakalamak için apne çıkışı
seslerini bulmayı amaçlamaktayız. Bu amaçla horlama kayıtları önce parçalara
ayrılmakta. Daha sonra bu parçalar literatürde bulunan ve bulunmayıp yeni
kazandırılan ses özellikleriyle ifade edilmekte. Son olarak da bu parçalar
denetlenmiş ve denetlenmemiş yöntemlerle sınıflara ayrıldı. Bulunan apne çıkışı
sesleri fazla bulunan parçaların sayısını azaltmaya yönelik geliştirilen bir yöntem ile
apne zamanlarının bulunması için kullanıldı. Aynı zamanda sınıflandırılan uyku
seslerinden doktorların faydalanması için olasılık hesabına dayalı bir kaynak
oluşturuldu. Bu yöntemler, Gülhane Askeri Tıp Akademisinde TUAS tedavisi için
ameliyat olacak hastaların ameliyat öncesi ve sonrası horlama sesi kayıtlarına
uygulanıp, tedavinin ne kadar başarılı olacağı belirlenmeye çalışılacaktır.

Anahtar kelimeler: Tıkayıcı uyku apnesi, apne çıkış sesi.

To My Family and Friends...

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Sleep is a block time period during which a person is not awake, although the brain continues to function [**1**]. It has stages; *rapid eye movement* (REM) and *non rapid eye movement* (NON-REM). REM is the stage that most of the dreaming occurs and NON-REM occurs before REM with four sub stages goes from first stage (light sleep) to fourth stage (deeper stages). As the sleep goes to a deeper stage, brain signals, eye movements, breathing, muscle conditions change. Sleep is routine of our body which is arranged by our biological clock and the quality of sleep is very important for our health, social life, success. There are many controlled tests applied to investigate the importance of sleep and sleep quality on our life [**1**,**2**,**3**]. Sleep disorders could disrupt the sleep causing cardiovascular problems depression, anxiety, panic attacks and in severe cases, even death.

As sleeping is a very big part of our life and our body continues working, we may face many disorders during sleep that requires a lot of attention. In *American Academy of Sleep Medicine,* the sleep related disorders are classified under eight main categories as shown in   Table 1.1 [**4**].

Table 1.1: Sleep Related Disorders

Sleep Related Disorders
Insomnias
Sleep-related breathing disorders
Hypersomnias Not Due to a Sleep-Related Breathing
Parasomnias
Sleep-Related Movement Disorders
Circadian Rhythm Sleep Disorders
Isolated Symptoms. Apparently Normal Variants and Unresolved Issues
Other Sleep Disorders

Snoring is the sound that is produced by primarily inspiration as a result of partial collapse of some parts of the upper airways at the back of the mouth or upper throat. This obstruction results in air turbulence, causing the tissues in the airway to vibrate and making the snoring sound [1]. It is a widespread problem mostly occurring in males, and is more serious than making an annoying sound during sleep. 20% of the men and 5% of the women who are 30-35 years old snore habitually and these ratio increase to 60% of men and 40% of women after the age of 60 [5]. It may be a symptom of many harmful diseases, which are usually confused with sleep disordered breathing. The most commonly encountered sleep disorders are apnea and hypopnea.

Hypopnea is the partial blockage of respiratory airflow due to partial obstruction of the soft tissues in the upper airway and mostly appears as a benign snore except that it causes reduction in the oxygen saturation in the blood [5]. *Apnea* is a Latin medical term which means *without breathing* [5]. Sleep apnea is the abnormal pauses in breathing (for more than 10 seconds) or moments of

abnormally low breathing during sleep due to obstruction in the upper airway or neurological defects [**4,5**].

Apnea and hypopnea are diseases that are not easy to detect as they happen during sleep. The most commonly used technique is keeping the patient in the sleep laboratory of a hospital or a sleep clinic and recording several biological signal with polysomnography (PSG). This technique is expensive, needs professionals to apply and analyze and also uncomfortable for patients. Due to these drawbacks, alternative diagnosis techniques have been investigated by many researchers. One of the most popular research areas for this goal is to gain information about apnea by using snoring sounds as they are strongly correlated.

During apnea period, as the breathing stops, no snoring sound is detected. Almost after every apnea, the silence period ends with a loud snoring noise due the opening of the airway. So those noises are indicators of apnea moments. This means that, if the post apnea sounds can be detected from the sleeping sound recordings, the apnea periods can also be detected without using PSG.

In very few studies, some of the acoustic characteristics of the post apnea sounds are mentioned to be different but not much attention is given that issue. In this study, we aim to show the differences of those loud noises, see if they are detectable and can be used for apnea detection. Also applying those results to clinical data is one of the major goals of our study.

1.1.   Contribution of the Thesis

With this thesis we aim to contribute apnea studies with the following:

- Using post apnea sounds in order to detect apnea periods
  - The characteristic of the post - apnea sounds are reported to be different but concrete representation of these differences have not been presented.
- Unsupervised and supervised grouping of various sleeping sounds.

- Introducing new characteristic differences between post apnea sounds and other sleeping sounds.
- Introducing new features to represent the differences between post apnea sounds and other sleeping sounds.
- Developing a sound based guide to help PSG analyzers in scoring the PSG recordings.

1.2. Scope of the Study

The following topics are covered in this study:

- Recording of the snoring sounds of OSA patients with non-contact microphones at the same time with PSG recordings
- Creating snore database
- Synchronizing the PSG recordings and sleeping sound recordings that are taken with two separate devices
- Labeling whole night sleep sounds with the aid of medical doctors by evaluating the PSG recordings
- Filtering the environmental and electronic noise for the sleeping sound recordings
- Activity detection in order to eliminate the silent parts
- Researching for and extracting proper features for each frame in an episode to evaluate sleeping sounds
- Constructing audio features that are not presented in the literature to evaluate sleeping sounds
- Using PCA analysis on the feature set to eliminate extra dimensions
- Extracting episode summaries for the feature vectors
- Using unsupervised classification algorithms to
  - group the sounds that have similar characteristics in feature based representation
  - observe the characteristic differences or similarities of predefined basic sound classes that are recorded during sleep

4

- o   show that post-apnea sounds are detectable

- •   Using supervised classification algorithms to
  - o   detect post apnea sounds
  - o   detect apnea periods
  - o   construct a numerical guide to aid medical doctors or PSG specialists in scoring the recordings

1.3.   Outline of the Thesis

As stated in before we basically work on acoustic properties of post apnea snoring sounds. The background information about snoring, its acoustic properties, apnea and its mechanism and previous studies about these two is represented in Chapter 2. After that what we use in our method to achieve our goal is mentioned in Chapter 3. How we construct a database, what kind of processes we do on that data base, what kind of acoustic properties and features are used in representing the sleeping sounds and the characteristic differences between post apnea and other sleeping sounds is described in this chapter. Also, the supervised and unsupervised classification methods and how we evaluate their results are subtopics of that chapter. In Chapter 4, we introduce our algorithm step by step with results, comments and discussions underneath them. An decision technique for undecided sleep sound episodes is given in the same chapter. At last, in Chapter 5, we conclude our study, present our observations and mention possible future work.

CHAPTER 2

BACKGROUND INFORMATION

In this chapter the background information on snore, its acoustic properties, definition of apnea and hypopnea will be given. Also previous studies will be examined in order to have a proper knowledge of what have been done and what is missing on the sleep sound investigation topic.

2.1.   Definition of snoring

As a dictionary definition, snoring is defined as "to breathe during sleep with harsh, snorting noises caused by vibration of the soft palate" [6]. The source of simple snore sound is the phryngeal segment of the upper airway. Although it is named as simple, it has mostly social side effects on a person's life. These side effects are hearing problems of the bed partner, social problems like anger due to decreased quality of sleep, etc. Divorces are also a significant social result of snoring [7,8].
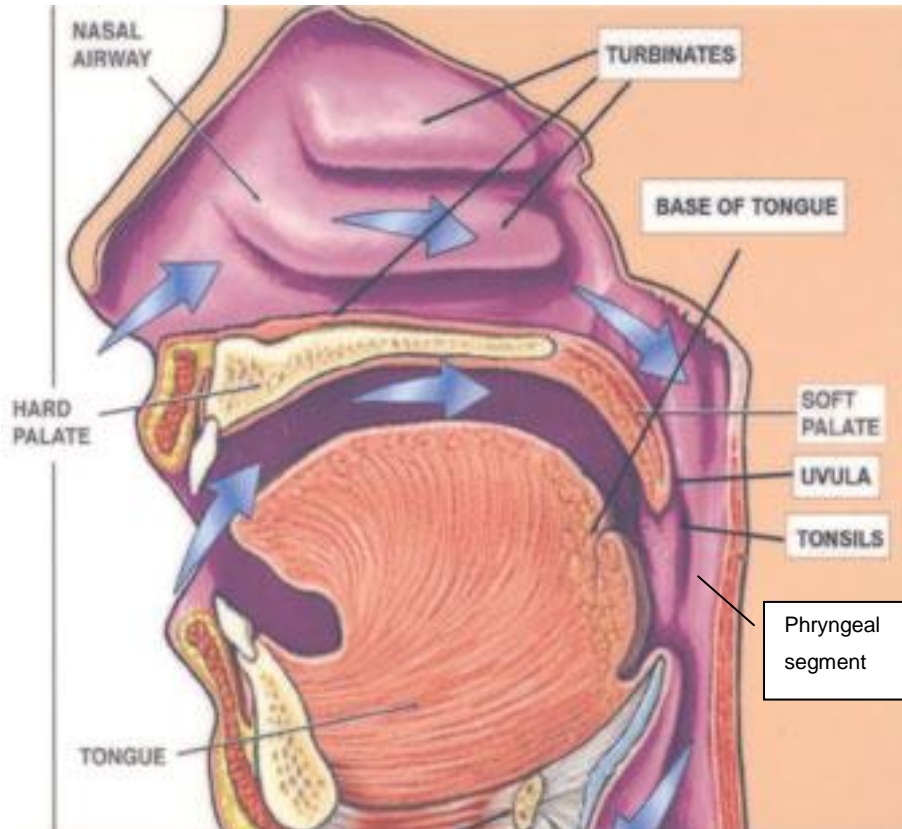
Figure 2.1: Upper airway [6]

2.1.1. Factors that Cause Snoring

Several anatomic factors are reported as causes of snoring. Insufficient muscle tone of the airway elements like palate, tongue or pharynx is reported as the most widespread cause of snoring [**5**]. In the deep stages of sleep those abnormal muscled airway parts fail to fulfill their job for the respiratory cycle and some abnormalities like unexpected position of the tongue. The tongue falls backward into the airway. This yields to vibration of the soft palate, uvula or pharyngeal folds seen in Figure 2.1 [**5**]**.** A receding chin, a nose structure that limits the airflow, excessive length of soft palate or uvula are some anatomic abnormalities that cause snoring. Existence of space occupying tissues is also reported as a biological factor that increases snoring [**5**].

Other than anatomic character of the humans the following are also mentioned as snore generating factors: Gender (due to laryngo-pharyngeal and body fat differences, males snore more than females), weight problems, smoking, alcohol/tranquilizers (due to relaxation of upper airway muscles), nasal obstruction, daytime sleepiness, witnessed apneas, abnormal sleep/night activities (sleepwalking, periodic movements, night sweats, reduced libido) [**9**].

2.1.2. How Snoring Sound is Generated

During normal breathing, air passes through the airway in Figure 2.1. Normally it causes breathing sounds that are hardly heard as air moves steadily without any obstruction. During sleep, due to narrowing of the airway, the airflow vibrates the unsupported soft parts of the upper airway, which generates snoring sounds [**5**].

2.1.3.Significance of Snoring

Although snoring is generally known as an annoying or funny sound produced during the sleep it is generally a sign of abnormality in the human airway. For what reason it is an indicator of the problems in the upper airway in any level [**6**]. It is reported that loud continuous or irregular snoring are frequently seen in sleep related breathing problems due to the obstruction of the upper airway [**10**]. It can carry information about the level and place of obstruction of the upper airway[**6**].

2.2.   Acoustics of Snoring

Snoring is mostly a low frequency harmonic noise. The acoustic characteristic of snoring sounds change due to many factors such as the route of breathing

(nasal, oral, both nasal and oral), the source of the snore (palatal snores, tongue based snores, supraglottic space based snores, or any combination of those), existence of sleep disorders like hypopnea or apnea, body position (supine, left, right), body movements during sleep, sleep stage, age, sex, weight of the patient, the shape of the airway [**5**,**6**]. Tracking the changes in the sound characteristic may supply information about the existence and significance of those factors. The fact that a person is a longstanding, habitual, continuous or interrupted snorer, that the snoring sound is loud or carrying a weak characteristic, duration and regularity of the snore are all important things to be considered by the medical doctors in order to catch abnormalities or diagnose a disease.

Snoring sounds are very similar to speech as they are generated by a mechanism similar to that of speech sounds. Upper airway acts like an acoustic filter for both. Also, they are both originated from vocal tract. Therefore, many speech recognition techniques are applied on the snoring sounds to investigate the properties of snore sounds and to gather information about the snore sound characteristics. There are some main differences between physiological characteristics of snore sounds and speech sounds; first, snoring is created by vibration of pharyngeal structures whereas the speech sound is created by vocal cords. Also, snoring sounds are produced while the upper airway is in a passive stage (sleeping), and no junctions or articulation of the sounds are observed. Furthermore, the forcing pressure to create the snore sound is mainly inwards as snoring sound occurs mostly during inspiration. These physiological changes result in some sound characteristic differences between speech and snoring sounds, which should be kept in mind while speech processing techniques are used for investigating snoring sounds [**6**]. First, speech signals vary in a very wide frequency band however snoring sounds are mostly located on lower frequency bands. It is also important to keep in mind that, snoring sounds have more correlation in a short time period than speech [**11**].

As stated before snoring sound can be used as an indicator of many sleep disorders, used for retaining some anatomic information especially about the upper airway. Beck *et all.* discovered the acoustic properties of snore sounds in both time and frequency domains. They observed sound sequences at 62-136

9

Hz as snoring sounds with rapid oscillations. They used three experimental setups: A dog model which an artificial upper airway obstruction is created, simulated snores and snore sounds from sleeping normal but heavy snorers. They examined total 120 snore episode taken from 6 dogs. The snore sounds are same shaped, have same amplitudes and frequencies. They call complex wave snores which appear as 3-8 peaks in 220-850 Hz frequency range in their power spectrum and result from colliding of the airway walls and represent airway closure. The simple wave snores are more like a sinusoidal wave where no secondary oscillation appears with in a periodic wave. They have peaks in 174-292 Hz range. Smaller peaks are followed by main peaks. They claimed that majority of snores can be classified with these two categories. Others have characteristic which are very hard to classify with closely placed peaks at their power spectra. From the second experimental setups they observed different shaped complex wave snores where the highest sound deflection detected when maximum oscillation in respiratory signal occurs (Figure 2.2) [**12**].
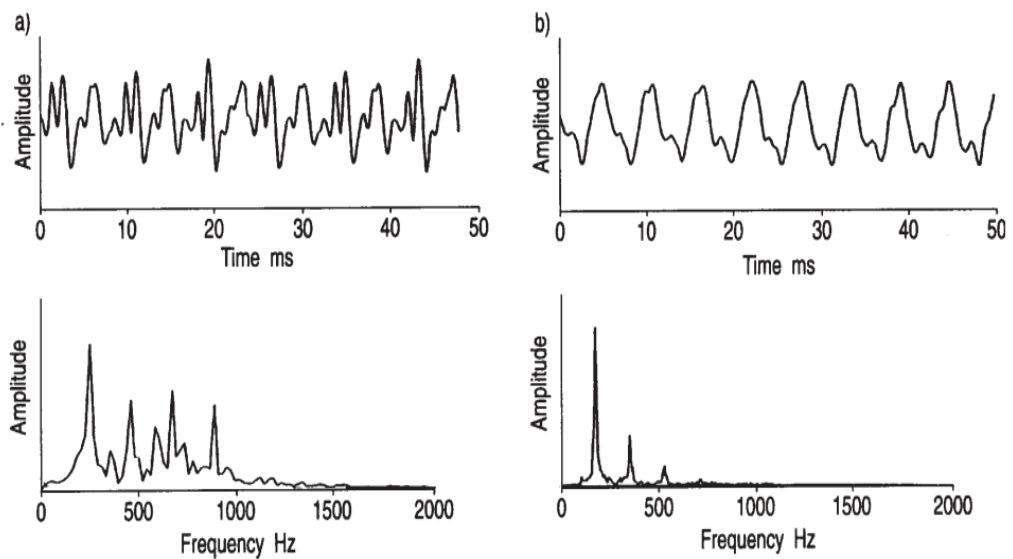
Figure 2.2: Difference between simple snore waves and complex snore waves, a) complex snore wave, b) simple snore wave [12**]**

Osborne *et all.* constructed a snore an acoustic model to differentiate the palatal and non-palatal snores. Their claim was palatal snores have a peak at about 20

Hz in their power spectrum which is a very low frequency. Considering this property they constructed a peak ratio which is a ratio of the 99% of the spectrum and the R.M.S. value of the spectrum. Using a threshold they were able to separate palatal and non-palatal snores with a very high success[13].

Dalmasso et all. worked on the acoustic characteristic differences in snore sounds in order to show the fact that different anatomic characteristics result with different acoustically different snore sounds. Suggesting three main audio features in order to evaluate the snore sounds belonging to different patients or snorers. One of those features is *Leq-Equivalent Continuous Sound Level* which is defined as the energy mean of the sound level averaged over the examined period. (2.1).

$$\text{Leq} = 10 \log_{10} \frac{1}{t} \int_0^{\tau} \left(\frac{P_A(t)}{Po} \right)^2 dt \tag{2.1}$$

Here $P_A(t)$ is the instantaneous A weighted acoustic pressure and P_O is the reference acoustic pressure. This energy based feature showed promising results [14,15].

By examining the frequency spectrum of the snore episodes it is possible to detect the source of the sound. After using a *hamming* window and taking the Fourier Transform of the episodes from different sources taken from 16 patients Agrawal *et al.* [16] discovered that the snorers that the tongue based snores have the peak frequency at 1243 Hz. which differs significantly from the other snore sources. The centre frequency of the tongue based snore, 1094 Hz, also supports the same significant difference. Other sources show different frequency characteristics as listed

Table 2.1 but not as distinct as the tongue based snores.

Table 2.1: Frequency characteristics of snores from different sources [16]

Subjects	No. of snores	Nasendoscopic evaluation	Peak frequency (Hz)	Centre frequency (Hz)	Power ratio
1–12	118	Palatal	137 (105, 189)	391 (253, 1027)	7 (2, 63)
1	10	Tongue base	1243 (1215, 1277)	1094 (1059, 1200)	0.2 (0.1, 0.3)
13	6	Mixed	190 (115, 223)	404 (312, 605)	5 (2, 50)
3	5	Epiglottic	490 (331, 510)	442 (393, 464)	53 (27, 84)
14,15	22	Tonsillar	170 (85, 201)	445 (332, 601)	12 (7, 17)
9	4	Secretions	116 (111, 137)	510 (359, 618)	2 (1, 6)
9	4	Unknown	161 (122, 230)	525 (501, 550)	2 (2, 3)

A similar study is carried by Haries's. They investigated if snoring sounds could be an alternative to nasendoscopy, which is a method applied under anesthesia, and used the snoring sounds to understand the site of snoring or cause of breath holding. They showed that the frequency distributions of tongue based snoring and multi-segmental snoring are different. They could not obtain the same success in differentiating the palatal snoring and multi-segmental snoring [17].

The strength of the airflow, vibrations of the soft palate, shape of the upper airway, the obstruction due to tongue position are all factors that affect the snoring sound. Because of those unrelated factors, in some studies the non-linear properties of snore sounds have been examined by some researchers. Mikami *et al.* investigated the nonlinear properties of snoring sounds in order to get information about the differences between normal snorers' episodes and the ones from the apnea patients. They verified the existence of non-linearity in the snore sounds by applying surrogate analysis and correlation integral methods. Up to their study, non-linear methods had not been studied for OSA diagnosis purposes. As it was explained before, the generation of snoring sounds depends on many different biological parts and their characteristics. Linear description of those factors is quite unrealistic. They observed that there are 3 or 4 stationary parts on the average in a single snoring episode taken from a normal snorer. In

their study those stationary parts are transformed into frequency domain and grouped under three main headers: simple, complex and harmonic [18].

2.3. Sleep Apnea

During the onset of sleep and while going deeper sleep stages there becomes some physiological changes in the ventilation system elements which causes sleep-disordered breathing. A reduction in tidal volume, an increase in upper airway resistance, reduction in the muscle tone of phryngeal dilator muscles during sleep and collapse of airway which effects patency are some of the changes effects the gas exchange routine and quality [19]. The most well known disease app ear with the above sympthoms are hypopnea and apnea.

Hypopnea can be defined as a reduction of airflow greater than 30 - 40% or clear reduction of airflow at nasal cannula pressure transducer which occurs of at least %3 decrease in O_2 saturation or may be observed from EEG signals [19].

Apnea is mainly grouped under two cathegories which are *central apnea (CSA)*, *obstructive sleep apnea (OSA)* and *mixed apnea* which is combination of OSA and CSA. A less prevalent version of obstruction in breathing during sleep; *central apnea (CSA);* is a disorder of decreased breathing rate or depth, particularly during sleep due to a transient reduction or withdrawal of central neural output to diaphragm and intercostals muscles. On the contrary that OSA is a physiological reasoned disorder, CSA is neurological based. In CSA the airway is not obstructed, but the breathing reflex is periodically interrupted, so there are no chest and abdomen reactive moments [19]. *Obstructive sleep apnea* (OSA), on the other hand, is a physiological-based disorder, in which the breathing stops for at least 10 seconds due to an obstruction in the upper airways. CSA is mainly shown as a symptom of heart failure. People with heart failure mostly have apnea syndrome according to studies, and CSA is shown as a symptom of heart failure as well as OSA, since the patient suffering from sleep apnea is more likely to develop abnormal and dangerous heart rhythms [19].

CSA is mainly classified under two headings: hypercapnic and nonhypercapnic central apnea [**4**].

OSA is the more common version of sleep apnea, which we will be dealing with for the rest of the study.

2.3.1. Obstructive sleep apnea

Obstructive sleep apnea is a sleep disorder disease that appears as the pauses in breathing due to closure of the airway. The obstructed part of the upper airway is shown in Figure 2.3. It happens several times during sleeping. The regularity of sleep apnea is a measurement criteria for the severity of the disease. It is called apnea / hypopnea index (AHI), which gives the number of apneas in one hour period.
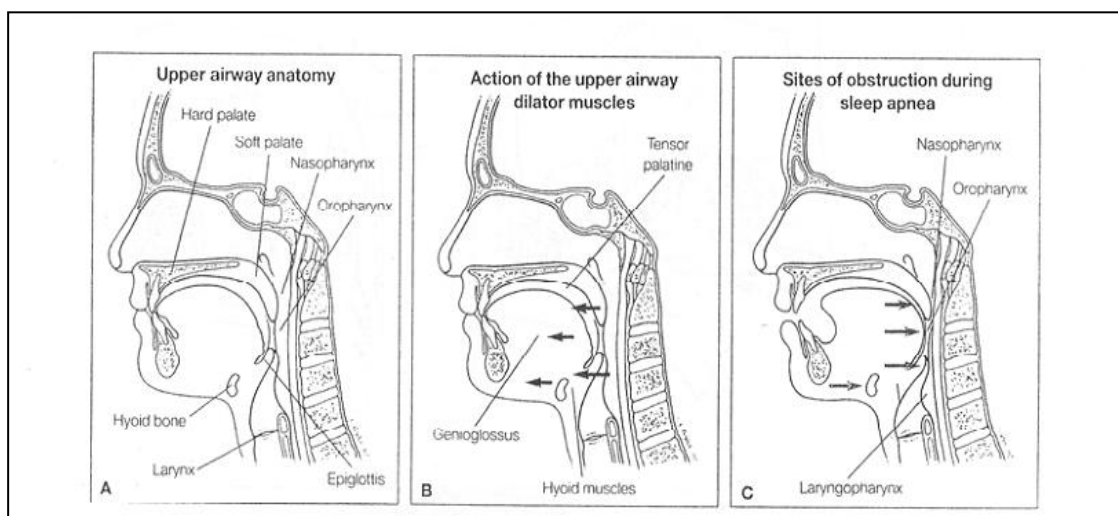
Figure 2.3: Anatomic view that shows the collapsible part of the upper airway [5**]**

OSA is seen among 2% of woman population and 4% of men population in Turkey. OSA is a serious health problem; it can cause cardiovascular problems such as increase in blood pressure (hypertension), strain in the cardiovascular system due to the decrease in blood oxygen level. These results can lead to

heart attack and even death [**4**]. Also daytime fatigue and complications with medications and surgeries are well known results of sleep apnea. Furthermore, it is reported in Pizza's study that severe OSA significantly effects the driving performance of the patients. From the experiments applied to over 24 male patients whose age is around 54, it was observed that there is a strong correlation between bad driving performance (simulated driving performance) and daytime sleepiness as a consequence of OSA. The criteria for bad driving performance were lane-position variability and crash occurrence [**20**]. Also heavy snoring, learning problems, excessive daytime sleepiness are detected in children suffering from OSA with ages between 6-11 [**21**]**.**

2.3.2. Diagnosis of Obstructive Sleep Apnea

OSA is a disease which is difficult to diagnose as it is happening during sleep and it is usually mixed with heavy snoring. There are some symptoms of OSA are listed in Table 2.2.

Table 2.2: Symptoms of OSA

Symptoms of OSA

o Breathing pauses during sleep

o Snoring

o Nasal Voice

o Noisy Breathing

o Sweating

o Mouth breathing

o Morning headaches

o Difficulty in waking up

o Restless sleep

OSA patients are mostly not aware of their disease. The diagnostic procedure needs a full-night sleep study with the most popular method polysmonography (PSG) and an evaluation of the results by professionally trained medical doctors, spending many hours to decide the existence and level of OSA. The PSG needs to be applied by a technician, as there are many sensors to be attached properly for measuring electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (EEG), chin electromyogram (EMG), nasal airflow, oral airflow, thoracic movement, body position, abdominal movement, blood oxygen saturation (SaO_2), leg electromyogram and snoring sounds [5]. EEG, EOG, and chin EMG provide information about the sleep stage of the patient. The ECG is used to observe cardiac rhythm or any abnormality if exists. To measure the airflows in nasal and oral parts, pressure sensors and thermistors are used. During respiration, the temperature due to the inhaled and exhaled air changes in our nose and mouth, the thermistor is used to detect those changes and give idea about any flow limitation that occurs during sleep. There are also sensors to measure chest movements to provide information about pause or decrease in respiration. The pressure sensor, thermistor and chest movement measurement

16

results are used to detect apnea / hypopneas. Body position sensor and a snoring microphone are attached to the patient to provide extra information during any abnormalities that occur during sleep. Also a *continuous positive air pressure (CPAP)* sensor is used for balancing the pressure in the airway in some cases. It is for both diagnosing the airway obstructions by measuring the pressure changes and temporarily opening the airway when the obstruction happens [**5**].

Although the PSG method is the most widely used OSA diagnosis technique, it has many drawbacks; it is an expensive method and not available in most of the hospitals. Besides, the patients are attached to many sensors and forced to stay at the hospital, and to sleep in an unfamiliar environment for at least 1-2 nights. Furthermore, PSG needs specialists to evaluate the results.

Figure 2.4: Patient going under a sleep test

Those difficulties in the PSG test oblige researchers to find an alternative method for diagnosing OSA. In recent years acoustic analysis of snoring became the most popular topic for this goal. Snoring analysis is a simpler technique to apply for both patients and doctors. Recording of sleeping sounds is a simple technique that requires microphones hung above patient's head. This also allows to make the recording of sleep sounds at the patient's home, in his/her own sleeping environment. The patients sleep better in their home environments, there is no extra financial loss for occupying hospital rooms, there are no attached sensors to the body that significantly ruin sleep quality [9].

2.4. Characteristic of Snoring Sound for Normal Snorers and OSA Patients

Many recent studies proved that the snore sound characteristic differs between the normal snorers and OSA patients. These studies will be discussed in the following part. The difference of the characteristic is mainly because of the changes like narrowing, tissue differences in the source of the snoring signal, etc.

The main concern is the differences in the power spectra of the snore episodes. When comparing the post-apnea snore episodes and normal episodes shown in Figure 2.5, significant differences are observed in the power spectra; in Figure 2.5-a the two normal snore episodes have more regular harmonic characteristic with a fundamental frequency of 100-150 Hz. As we investigate the two post apnea episodes given in Figure 2.5-b one can easily see that the regularity of the power spectrum significantly vanished and talking about a periodicity is nearly impossible [24].

Similarly Perez *et al.* examined 10 apnea and 10 non-apnea snore episodes, and compared their power spectra. What they concluded was most of the power is concentrated below 2000Hz, and the peaks are mostly below 500Hz for simple snores. For the post apneic snore sounds taken from the OSA patients, they

reported a more broad band white noise which have more power at the higher frequencies [**22**].
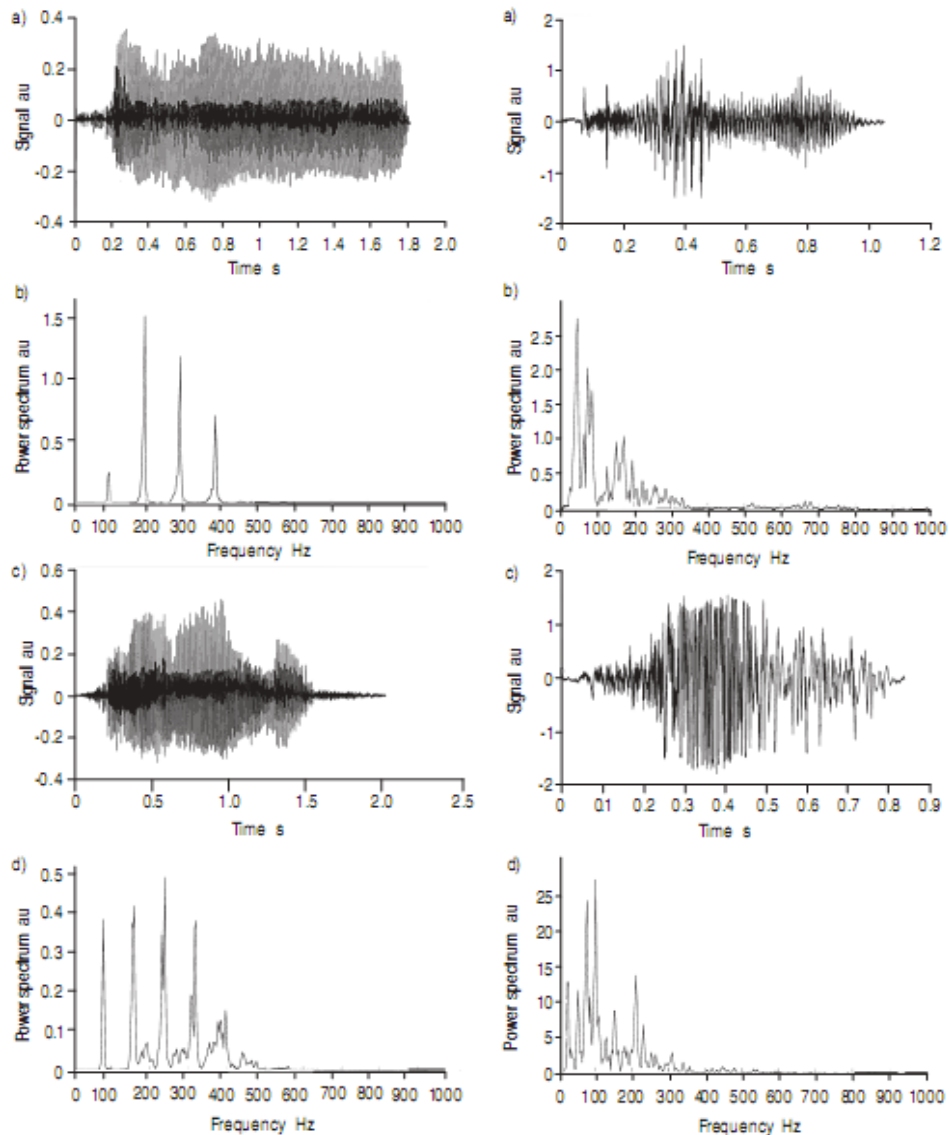
Figure 2.5: Differences in the power spectrum of the snoring episodes of a) time domain differences between normal snorers and apnea snorers b, d) frequency domain differences between normal snorers and apnea snorers [22]

Cavusoglu *et all.* studied the sequential properties of snore signals in order to find a diagnostic approach for OSA from these signals. They basically aim to detect the characteristic of the snore by using snore regularity. The properties that they used in this study are, snoring episode durations (SED), snoring

episode separations (SES) and average snoring episode powers (SEP). They showed that that these sequential properties show a higher variation in the patients with OSA than the normal snorers. Investigating the mean, standard deviation and coefficient of variation of those features, spider charts were composed to visualize the differences between the joint distributions of the statistical variables of OSA snorers episodes and benign snorers [23]. They proved that snoring signal can be a sign of OSA.

Table 2.3: Statistical comparison of SED [23]

SED parameters	Simple snorer	OSA patient
Average SED (s)	0.77	0.72
Maximum SED (s)	2.15	3.35
Coefficient of variation of SED	0.35	0.57

When patients enter to an obstructive apnea period we expect to observe a quite or occlusion voices like choking and at the end of they leave the apnea period (start breathing again) with a loud voice which has a different characteristic from other remaining snoring episodes. It is claimed that that loud post-apneic snore consists of mainly broad-band noise and has more power at higher frequencies [22]. But this property is not used in the literature in detecting the instants of apnea, or any other type detection mechanism.

2.5. Review of the Literature on Diagnosing OSA from Snore Sounds

Diagnosis of sleep apnea is a very popular topic and in recent years researchers worked on this topic deeply.

Many authors worked on characterizing snore according to its acoustical properties. Also automatic detection and classification of the snoring sounds have been a research topic in recent years. Duckitt *et al.* classified sleeping sounds by using speech recognition techniques. They used an HMM based sound classifier based on first 13 MFCC coefficients. With this method they were able to separate snore sound, duvet sound, breathing, silence and other sounds by using a small training set. The success of this study is limited to a little variety of data and to small data set which leaves it as improvable [24].

K. Ng *et al.* [25] assumed that snoring mechanism is just like speech mechanism due to the similarity that upper air way acts like an acoustic filter in producing snoring sounds and vocal tract acts in the same manner as required for speech signal. So they focused on the well known speech processing tool, LPC which is a technique for modeling vocal tract. LPC gives formant information about snore, which is the information about the resonances that occur in the upper airway. The data-set containing 130 snoring episodes showed that there are differences in the first three formant frequencies when normal snorers and the OSA sufferers are compared. This result is shown by Receiver Operating Characteristic (ROC). The threshold for the first formant frequency to distinguish apnea and non-apnea snore episodes is defined as 720 Hz. The episodes that have first formants below this value are found to be taken from benign snorers and above are assumed to be taken from the OSA patients. This research may be successful in showing the difference between the OSA snorers and benign snorers but experiments should be performed for a lot more than 130 episodes to accept 720 Hz. threshold as a universal truth. Modified Normalized Mean Square (MNLMS) technique is used to reduce the noise coming from air-conditioner and other background noise present in the recording environment. One of the two non-contact microphones was placed as a reference microphone to get the acoustical noise in the area. Subtracting this reference input from the total input and applying the result to an adaptive filter an overall 6 to 9 dB noise cancellation was achieved.

Abeyratne and his group, separated the snoring signal in to three parts as breathing, snoring and background noise [26]. Then they further separated the snoring part as voiced and unvoiced (breathing). Zero crossing and energy are

the basic features used in this part. They also introduced a feature called intra snore pitch jump (ISPJ), which is used as a diagnostic sign of OSA [**26**].

Hara *et all.* proposed a multidimensional voice program to detect the differences between the simple snorers and OSA patients [**27**]. Standing on the truth that snoring is not a monotonous sound and actually a noise caused by vibration of the walls of orpharynx, they examined soft Phonation Index (SPI), peak frequency, Noise to Harmonics Ratio (NHR) and power ratio via Mann-Whitney test. SPI is a ratio of the lower frequency energy and higher frequency energy. The low frequency is defined to be in between 70-1600 Hz and high frequency is defined as to be in between 1600-4500 Hz. NHR is the ratio of the average spectral energies of harmonics in between the frequencies 1500-4500 Hz and harmonics in between 70-4500 Hz. Extracting those features of 48 male and 10 female patients snore episodes (30 episodes from each patient) They observed significant differences in those values as a result of the Mann-Whitney test. According to the study NHR and SPI gave good results in discriminating OSA and normal snores [**27**]. But they arranged the frequency ranges according to speech analysis so further analysis should be done by arranging the frequency ranges to more appropriate ranges for snore analysis or by adding more features to make the results more trustable. Also 30 snore episodes from each patient is a quite low value in order to investigate a whole characteristic of apnea and non-apnea snores.

Liao *et all.* worked on classifying sleeping sound episodes as snore and non-snore taken from all-night sleeping data. In order to separate the snore sounds as OSA and normal, they claimed that loudness of the snoring sounds that are produced by the OSA patients will be a discriminating factor for OSA and normal snore. Also, according to the study, deviation difference between OSA snore sound and normal snore is another identifier for OSA. Energy variation is used to identify the beginning and end points of sounds. Afterwards periodicity of the episodes is considered to differentiate the sounds as snore and non-snore. Then for further discrimination a set of features is used which include; fundamental frequency, ZCR, formant frequency, entropy of the spectrum and autocorrelation. KL difference and entropy of the spectrum is used to cluster the snoring sounds. Although their success seems considerable as they obtained 96.6% accuracy in

classifying the episodes as snore and non-snore sound and 94% accuracy in discrimination OSA episodes from non-OSA episodes, their data consists of only one person's sleep sounds recorded for only one night which may not be sufficient to be sure about the accuracy. Besides those reported differences may be associated with body signals that give information about sleep stages and body position may be a further improvement to this study [**11**].

Another similar study is done by Abeyratne and his friends. Again they separated the snore sound into snoring, breathing and silence by using log-energy, ZCR, normalized autocorrelation coefficients and first of linear prediction coefficients as features derived from the sound. They used a new pattern recognition technique to separate the sounds in to 3 sub-classes. The pattern recognition considers the mean and the variance of the features and classify them according to a training set. The success of the algorithm is around 90% however with different noise reduction algorithms it is increased up to 96.78%. The noise reduction techniques used in this study are: amplitude spectral subtraction (ASS) power spectral subtraction (PSS), short time spectral amplitude (STSA) [**28**].

Cavusoglu *et all.* They used sub band energies and ZCR, which are calculated from 100 ms windows that overlap by 50 ms. PCA analysis is used for dimension reduction in the feature vectors. The resultant feature vectors are separated with robust linear regression technique. 97.3% accuracy has been achieved in separation of the snore sound from other sounds in simple snorers however the accuracy value drops to 86.3% for the snore signals taken from OSA patients. Detection of snore episode will be useful in diagnosing OSA in terms of snore duration length, snore regularity, snore to sleep ratio and similar variables. This algorithm's performance may not be this successful under noisy conditions [**29**].

Wavelet analysis is used in detecting OSA by Strang *et al.* When compared with Fourier analysis, that method outshines to represent the discontinuities and significant peaks. Also it has multi-resolution analysis and synthesis ability [**30**].

As it can be seen from the above literature review, the acoustic knowledge of snore sounds is very valuable in detecting apnea. Among all these valuable

studies, the information supplied in [**22**] became our key idea, using post-apnea. With the prior knowledge of power spectrum differences between post apnea snore episodes and normal snore episodes, we continued our studies on post apnea episode discrimination and detection. Also, the study about classification of all night sleep studies [**11**] inspired us. As for our goal we tried to classify the sleeping sounds and detect post apnea episodes if they are detectable. Following these, we tried to use the information that can be obtained from post apnea episodes, in apnea period detection. Our method for these goals will be described in Chapter 3.

CHAPTER 3

METHOD

In this chapter the details of our method is described. To begin with investigating the acoustic properties of post apnea sounds, first we needed to obtain sleeping sound recordings including post apnea sounds and construct a database. The details of this step is described in the following part.

3.1. Recording Setup

All the experiments for this study has been done with the snoring data recorded from the OSA patients that are undergoing a sleep test at Gulhane Millitary Medical Academy (GMMA). The sleep test in GMMA consists of a full polysomnography. To get benefit from those data, our recording setup is arranged to take the recordings at the same time with polysomnography. The recording setup used to acquire sleeping sounds consists of *Studio Projects - C1* single diaphragm pressure gradient condenser microphones that have cardioid pattern and have a frequency response between 30 - 20000 Hz; and Edirol UA-1000 model multi-channel data acquisition system that is connected to a portable computer. The microphones were hung 20 cm above the patient's head. The data is digitized with 16000 Hz sampling rate and 16 bit solution. Initially, the recordings were taken by placing the setup in another room to avoid additional background noise that may be caused by the computer or the data acquisition system; however, this resulted in additional electronic noise due to very long recording cables. So all the system was placed in the same room where the patients were sleeping.

3.2. Snoring Database

We obtained 13 full night sleeping sound recording from GMMA sleeping centre. The recordings were taken from the patients that appeal GMMA with the symptoms of OSA. All of the patients are male and over the age of 50. After synchronizing the recordings a detailed annotation is done with the medical doctors in GMMA. Looking from the EEG, flow meter and thermistor outputs medical doctors label the time instants that the patient snores normally or goes to hypopnea or apnea periods. A sample of PSG recordings for apnea and hypopnea instances are shown in Figure 3.1 and Figure 3.2 respectively. The ECG signal, EEG signal, O_2 saturation, voice taken from a contact microphone, the flow meter signal, thorax movement signal are seen on the PSG screen shots in these figures. Looking those signals, they label the time instants of beginning and ending points of those periods. Finding those time instants from our sleep sound recordings we are able to find those special sounds to create our training and testing sets. As this process takes several hours and because of those doctors long working hours, we were able to annotate only 7 of the recordings up to now.
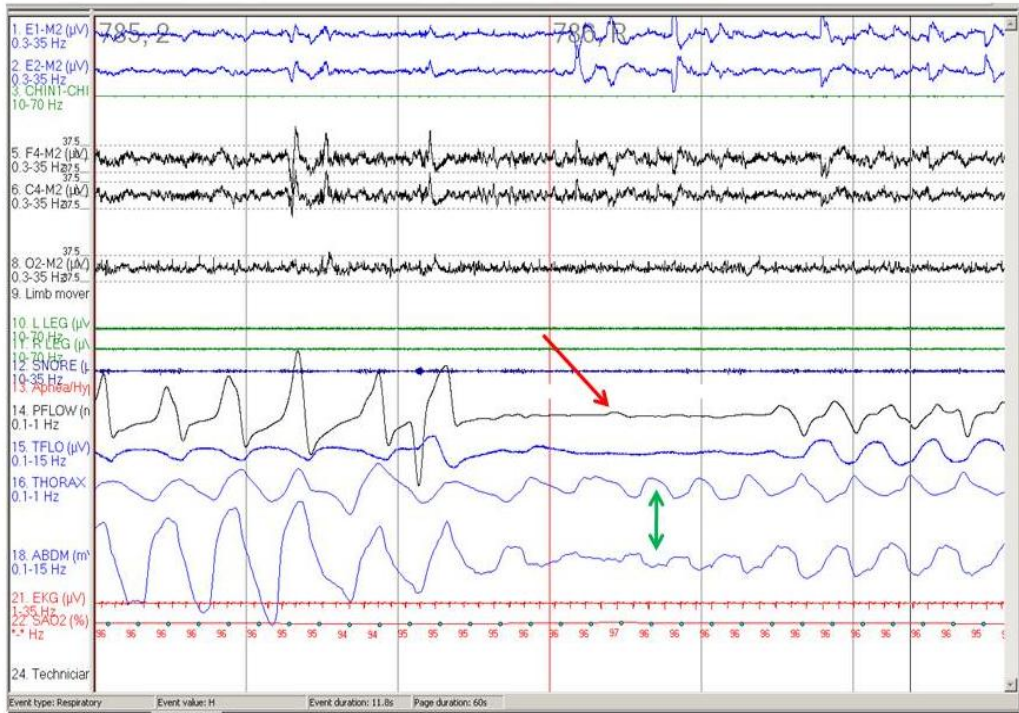
Figure 3.1: PSG recording sample showing OSA; pause in flow meter (red arrow) despite the respiratory effort (green arrow) [31**]**
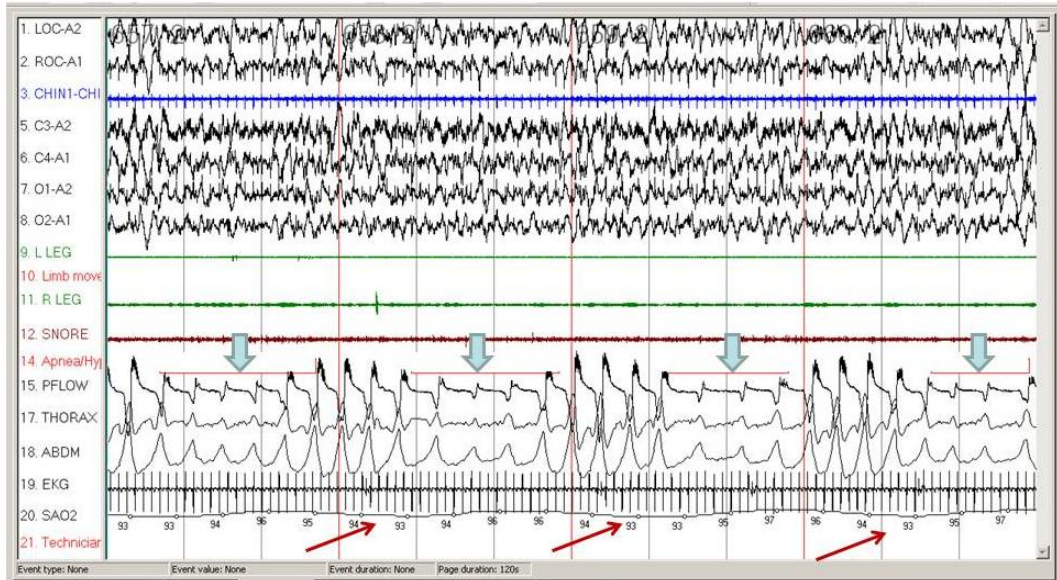
Figure 3.2: PSG recording sample showing hypopnea; arrows showing the oxygen desaturations [31**]**

The episodes in the recordings are grouped into 5 subcategories with the guidance of the information gathered from PSG. These subcategories are: normal snoring episodes, hypopnea period snoring sounds, post hypopnea snoring sounds that occur at the end of the hypopnea period, post apnea snoring sounds and the unexpected sounds occurring during apnea period. Randomly selected four episodes from each group is placed one after another in Figure 3.3. In the upper panel the time domain characteristics of the samples for each subgroup are shown. In the lower panels the frequency characteristics of the samples for each subgroup are shown.
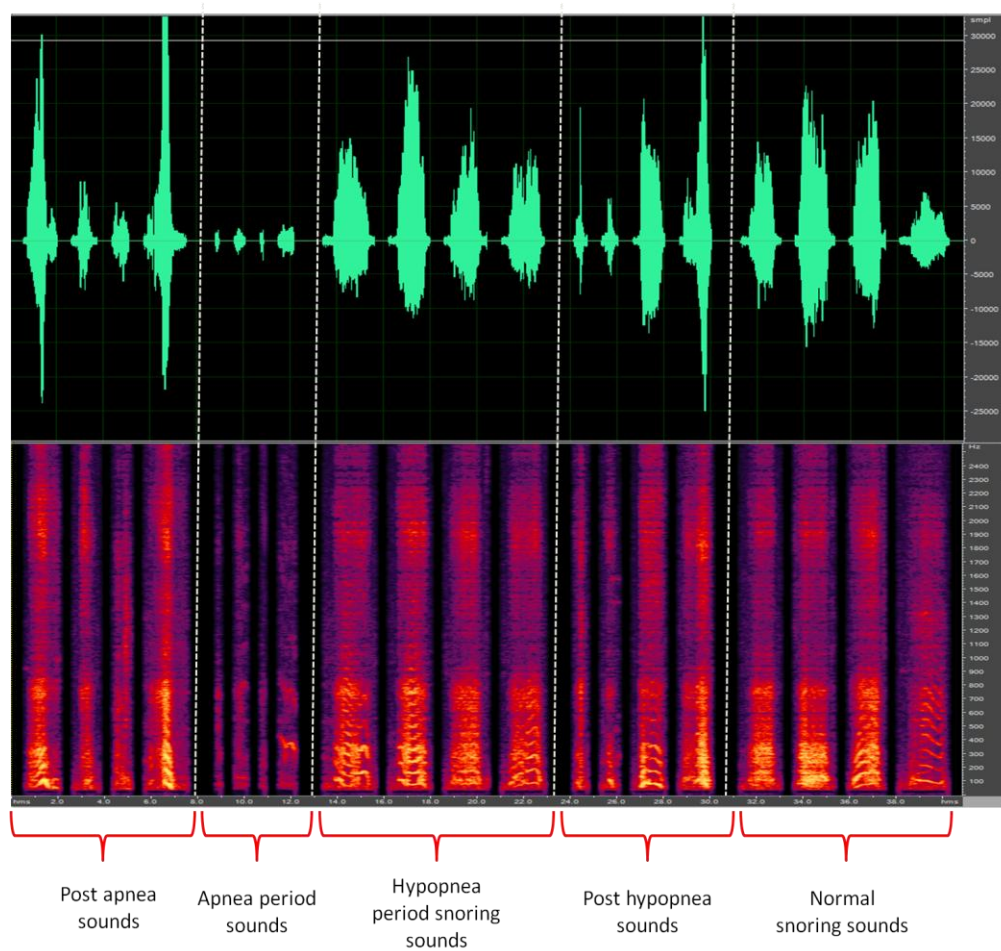
Figure 3.3: Four sample episodes from 5 different subgroups

3.3.   Pre-processing of Snoring Sounds

In the recording setup, sensitive microphones are needed so that any sleeping sounds are not missed. But high sensitivity of the recording system brings a significant drawback, which is the low frequency electromagnetic and background noise that interfere with the sleeping sound recordings. Loud snoring episodes are not much affected from this noise, but as the snoring sound has a dynamic range of 90dB [31], there may be low intensity snoring episodes that may be masked by the mentioned noise.

First as a precaution, the cabling is shortened to avoid extra noise interference and the recording setup is placed close to the patient. For a further precaution, the high pass filtering property of the microphones is used with a cut-off frequency of 75 Hz. Examining the spectrogram of the recordings, it observed that further low frequency filtering is needed to eliminate the noise. So a high pass FIR filter is designed to further decrease the effects of odd low frequency noise. The higher frequency noises are of less concern, because they do not mask the episodes significantly, and filtering those would cause losing the relevant information about the episodes. For those cases, adaptive filtering with *Least Mean Square* method may be used.

For cancelling low frequency noise, FIR filters are implemented. Linear phase and stability properties of FIR filters are the most important reasons for using them. Also the analog filters in the microphones are IIR filters, and their effect seemed limited from our observations. Parks McClellan high pass FIR filter design method is used. Pass-band, stop-band frequencies, the filter order are determined experimentally by observing the requirements detected from the spectrogram results.

3.4.   Acoustic Features

In order to make a successful classification the sleeping sounds must be represented by appropriate features. At this step feature selection plays the most important role in classification because bad choice of features may result overlapping clusters or wrong and meaningless classification results. In this part the feature set that is used in the classification and grouping algorithms are explained.

3.4.1. Energy

The short time energy is an indicator of the signal strength [**33**]. For our case it will be used to show strength differences occur in different frames. It is basically the sum of magnitude square of each sample in a single frame normalized with the length of the frame:

$$E = \sum_{i=0}^{N-1}(x_i)^2$$

(3.1)

N: number of samples in a frame

Energy is the basic feature to be used in a segmentation system for selecting the voiced parts of the recordings. The segmentation algorithm and how short time energy is used in that algorithm will be explained in section 3.5. Also, during night, the strength of snoring episodes may vary due obstruction, sleep stage, sleep position, etc. It is reported by the medical doctors that the post-apnea snores tend to have more energy.

3.4.2. Zero Crossing Rate (ZCR)

ZCR is the total number of times that the audio signal crosses the zero line with in a frame [33]. It is can also be defined as the average number of sign changes in a frame [34]. Considering this definition, it is calculated as in:

$$ZCR = \frac{1}{N} \sum_{i=1}^{N-1} abs(sign(x_i) - sign(x_{i-1}))$$

(3.2)

$$N = \text{number of samples in a frame}$$

Generally, ZCR has been used as a feature to differentiate human and environmental sounds [11] [35]. There are voiced and unvoiced frames in whole night snoring sound records. In the voiced parts, the number of ZCR is lower while in the unvoiced parts it tends to have a larger value. Considering this difference, ZCR is thought to be a differentiating audio feature between different types of sleep sounds.

3.4.3. Spectral Roll-off Frequency

Spectral roll-off is a measure of skewedness of the spectral distribution that is obtained by finding the frequency by which 95% of the total energy of each frame is focused [35]:

$$\sum_{f \leq K} X(f) = 0.95 \sum_{f} X(f)$$

(3.3)

$$\text{"K" is the roll-off frequency}$$

Higher frequency components of the audio samples tend have higher roll-off frequency (right skewed) [36]. It is noted in the literature that apneic snores have

31

more energy in the higher frequency bands [**6**,**25**,**27**]. Also the post-apnea snores tend to have most of their energy in the higher frequency bands. Existence or degree of obstruction is a factor that effects the frequency of the sound as the shape of the airway and the characteristic of the source of the sound changes. Considering these observations and anatomic facts, spectral roll-off frequency is considered as a valuable feature in discriminating snoring sounds.

3.4.4. Spectral Flux

Spectral flux is a measure for how quickly the spectrum of the sound sample is changing. Its value can be evaluated by calculating the summation of differences between consecutive samples in a single frame [**36**].

$$SF = \sum_{i} \sum_{FFTsize} (|X(i)| - |X(i-1)|)^2$$
$$i: \text{frame number} \tag{3.4}$$

Spectral flux attains lower values for periodic sounds like normal snoring sounds. Also the pitch jump probability is found to be high in OSA patients' snoring episodes [**37**]. These pitch jumps are sudden large changes in the spectrum and it can be observed from the spectrogram plots of OSA patients snoring episodes. Thus, we include this feature in our feature set.

3.4.5. Spectral Centroid

Spectral Centroid represents the represents the balancing point of the spectral power distribution in a frame. In other words it is the center of mass of the spectrum. [**35**]

$$SC = \frac{\sum_{i=1}^{N} k \cdot FFT(x_i)}{\sum_{i=1}^{N} FFT(x_i)}$$

N = the number of samples in a single frame   (3.5)

K = 0.7

Similar to spectral roll-off frequency, spectral centroid can be used in situations where frequency localization is different between the sounds to be discriminated. The spectral centroid takes higher values in post-apnea sounds as most of the energy is localized in the higher frequency bands [22] in post apnea sounds, compared to the normal snoring sounds, which makes this feature selectable.

3.4.6. Spectral Entropy

Entropy is a very popular feature used in information theory as a measure of information, choice and uncertainty. For audio classification applications entropy can be used to give information about the *peakiness* of the spectrum. For voiced sounds, the histogram of the investigated audio sample will give us clear formants, which appear as dominant peaks in the spectrum of that sample. For those frames, the entropy will be low. For the unvoiced cases we expect to observe higher entropy values [20].

To obtain *Spectral Entropy* for each frame, first FFT is applied to those frames. Then the probability distribution function is calculated by normalizing the spectrum of the frame over all frequencies [38]:

$$pdf = \frac{FFT(x)_i}{\sum_{i=1}^{N} FFT(x)_i}$$

where,   (3.6)

N is the total number of points in the spectrum (order of the FFT)

After this step the negative spectral entropy is calculated with the following formula:

$$H = \sum_{i=1}^{N} x_i \cdot \log_2 x_i \tag{3.7}$$

This feature is used to discriminate human speech sounds and snoring sounds [11]. Existence of speech data in our recordings, made this feature selectable for our feature data set.

3.4.7. Energy Entropy

Energy entropy is a feature that characterizes the complexity of the audio signal with the changes in time domain. It is used to detect the abrupt changes in the energy level of the signal. To calculate energy entropy, first the short time energy is calculated as in (3.1). Then the entropy value of the energy is calculated as[39]:

$$EnEnt = \sum_{i=1}^{N} E_i \cdot \log_2 E_i \tag{3.8}$$

Energy entropy takes small values for the frames where large changes in the energy occurs. In the postapneic snoring episodes, we are expecting sudden changes in the energy due to the opening of the airway. Thus, we believe that energy entropy will be a successful feature in detecting those episodes

3.4.8. Initial, final points and time constant of the LP error curve

Linear Prediction is a widely used technique in speech processing that generally works for modeling a source filter for speech [20]. It is an *autoregressive* analysis for extracting the characteristic of glottal, vocal tract and lip radiation transfer functions H(z) as a one all pole filter to model acoustics of vocal tract [40]. The linear prediction model is obtained by predicting the coefficients of the all-pole filter given in (3.9) where p is the model order and $\hat{a}(i)$ are the model coefficients.

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} \hat{a}(i) z^{-i}}$$

(3.9)

Although some components of speech can be expressed with a pole-zero model (auto-regressive moving average model) due to formant structure of speech, all pole model is preferred. The coefficients are estimated using Levinson-Durbin method in MATLAB based on least squares information [41]. The aim in Levinson-Durbin method is to find the coefficients that minimizes the error:

$$e = [s(n) - s(n)\,']$$

(3.10)

The prediction process is shown in Figure 3.4
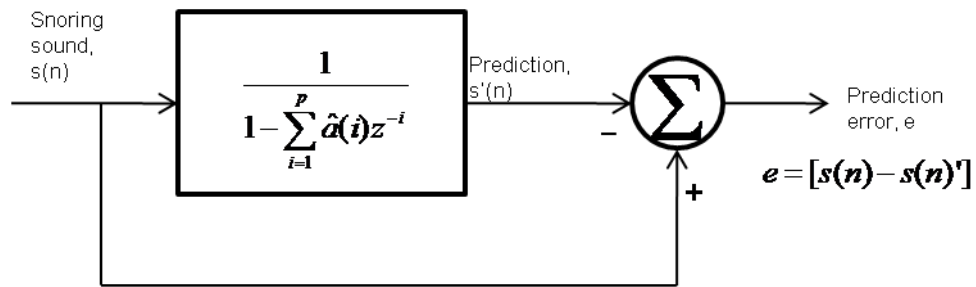
Figure 3.4: Linear prediction process

Since generation of mechanisms of snore and speech share similarities we focus on this feature to characterize the different snoring sounds.

The higher the order the of the LPC is, the more detail we can get about its spectrum as in Figure 3.5. We can assume the low ordered spectrum is the smoothed version of the higher ordered spectrum. This is because of the fact that, in the low ordered case, less number of poles are used to represent the LPC's where as in the higher order LPC spectrum more poles are used. So every detail can't be handle for very small orders. Also spectrum may miss in showing some harmonics in order to smooth the LPC spectrum as in Figure 3.5. For example the peak value showed by the red star on Figure 3.5,b is not observable in Figure 3.5,a.
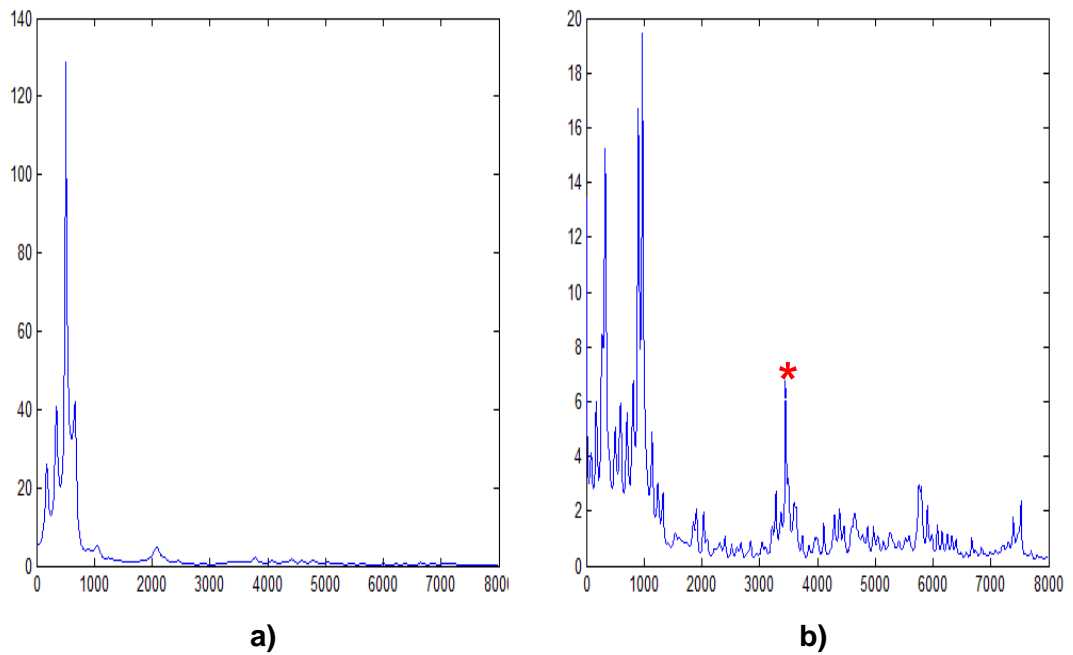
Figure 3.5: Difference between low (a) and high (b) ordered LPC spectrums

LPC error increases as the harmonicity of the waveform increases, because the spectrum includes sharp peaks that should be represented well. Smoothing the spectrum of harmonic data (*i.e.*, using a small number of coefficients to predict the spectrum) causes losses in the information so the predicted spectrum becomes different from the original one. As a result, the Figure 3.4 increases for the harmonic data represented with a small number of LPCs. Considering this fact we can say that the LPC error carries information about the harmonicty of the waveform. So the initial point (prediction with one coefficient), the final point (prediction with maximum number of coefficients) and the decay rate became important parameters in Figure 3.6 as they can supply information about the harmonicity of the waveform. We know that the post apnea sounds are much like white noise and the harmonicity is lower than the normal snoring episodes, so these new features will be good feature for discriminating post-apnea episodes from the normal snoring episodes. In Figure 3.6,a and Figure 3.7,b we can observe LPC prediction error curves taken extracted from different frames. The information from the curve can be extracted as follows:

The initial and final points are easily captured from the decrease curve. To obtain the decrease rate we first tried to fit the curve to an exponential decrease curve as they show very similar characteristics. The following expression must be solved in order to make the exponential curve fit.

$$\min_{\tau}(\sum_{k=1}^{n}((X_0-X_f)\cdot e^{-\tau_1 k}+X_f)-lpc(k)]^2) \qquad (3.11)$$

But the fitting was not so well as seen in Figure 3.6,b. Also $\frac{1}{k}$ decay is tried but again the fitting results gave high amount of error as in Figure 3.6,b. A more complex fitting equation is used as in 3.12 in order to obtain more precise fit. In this equation X_0 and X_f are known initial and final values for the LPC prediction error curve and what we need to find from this equation is the time constants τ_1 and τ_2 where one of them is for exponential term and the other is for the $\frac{1}{k}$ term.

$$\min_{\tau_1,\tau_2}(\sum_{k=1}^{n}[((X_0-X_f)\cdot e^{-\tau_1 k}\cdot(\frac{1}{k})^{\tau_2}+X_f)-lpc(k)]^2) \qquad (3.12)$$
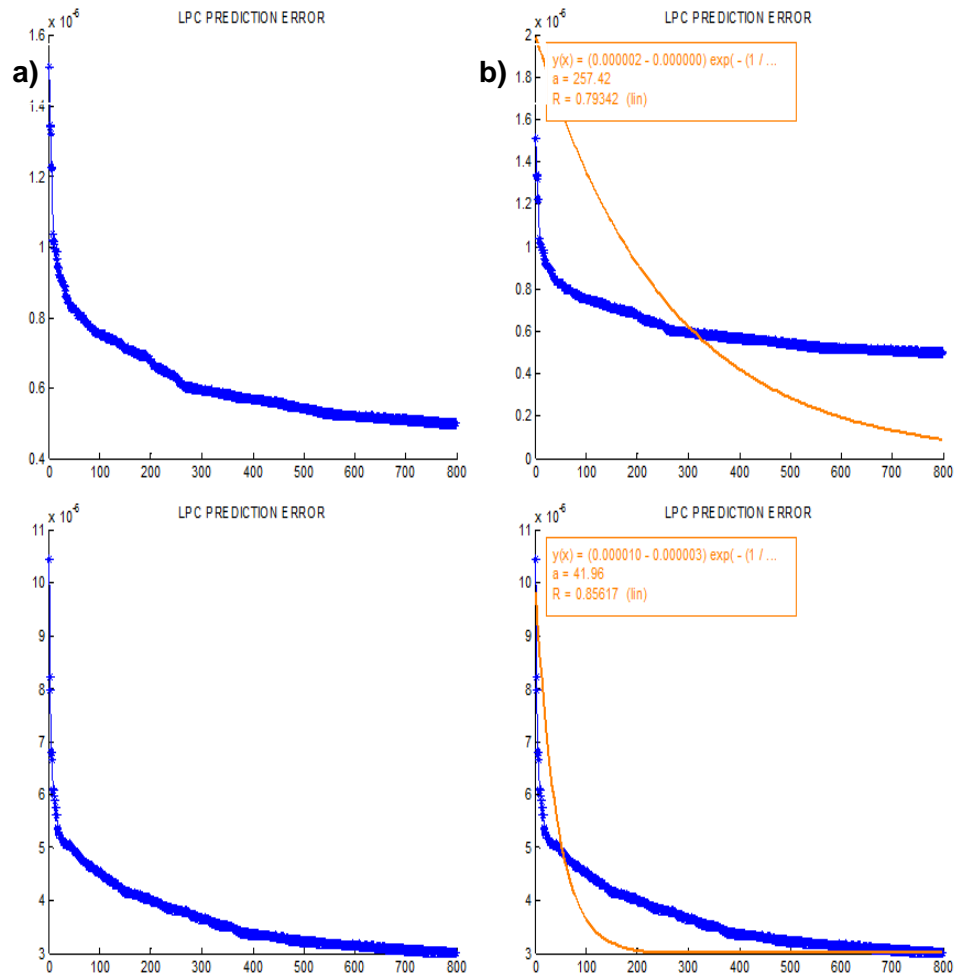
Figure 3.6: a) LPC error curves with increasing order, b) exponential fit on LPC error curve and related fit function chractersitics menttioned on the right
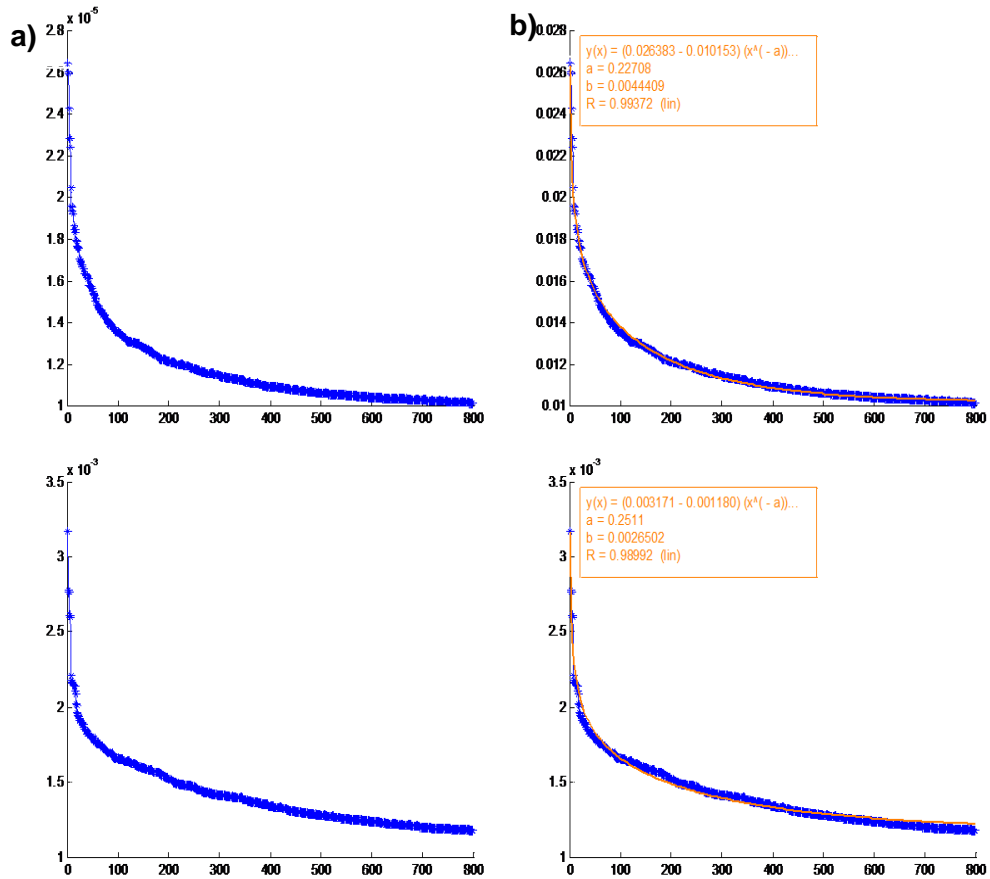
Figure 3.7: a) LPC error curves with increasing order, b) exp(x)*1/x fitted to LPC error curve and related fit function chractersitics menttioned on the right

The minimization is done by the Nelder-Mead Method [**42**,**43**] which is an unconstrained minimization method. As a result we obtain four new parameters that can be useful for classifying problems based on harmonicity.

3.4.9. Principle Component Analysis (PCA)

PCA is a technique to reduce the dimension of the feature vectors without losing much information [**44**]. This method basically finds correlated values of the feature vector and by applying some mathematical operations it reduces the number of dimensions, and converts the feature vector into a new vector consisting of uncorrelated values. Thus, mathematically it transforms the feature set to a new coordinate system where such that the highest variance of each dimension of the feature vector by any projection of the data comes on the first coordinate and the others compose the other coordinate axis with decreasing variances [**45**].

In order to obtain *principle components,* first the mean values of each dimension are subtracted from all of the data along the corresponding dimension to ensure that the first *principle component* describes the direction of maximum variance. The next step is extracting the covariance matrix, which contains scaled sums and cross-products of the feature values. The eigenvalues and eigenvectors are calculated from this covariance matrix. After sorting eigenvalues, the eigenvectors corresponding to highest eigenvalues are selected to be the new representation of the feature vector. Usually, the user selects the number of principle components to be used. However, this number can also be chosen by defining a threshold to eliminate the less significant components so that the dimension reduction can be done by avoiding much data loss. As the final step, the data is multiplied by the principle component vector so that a new feature vector is obtained.

3.5. Segmentation

In the scope of this study, snore episodes or other sleeping sounds are the parts that need to be focused on. However, silent parts in the recordings take almost the half of the recording time, and classification of complete signals increases computational load and causes unnecessary classes in the outputs. Manually selecting sound episodes takes a lot of time and may not always be robust, as it is a user dependent task; the boundaries between the silent and active parts may slightly differ from user to user, which may affect the results. Thus, to overcome these problems, we aimed to apply a segmentation method to eliminate the silent parts between the episodes before the feature extraction in some applications.

Cavusoglu *et al.* also used ZCR and short time energy to detect episodes [**29**]. In Cavusoglu's study, after calculating these features, a threshold mechanism is composed with user defined weights and decision of active parts is given according to this [**29**]. For our problem we use a similar approach. Short time energy is also used in our segmentation system but instead of ZCR we used spectral flux as it seemed to differ significantly between silence (only background noise exists) and active parts as spectral flux is a measure of spectral changes. The spectral change in silent parts and the active parts is absolutely different. In some cases only short time energy may be enough for detecting the episodes because the energy of the active parts is always higher. A simple threshold may be enough for simple cases where no significant background noise exists. But, as the recordings include background noise, additional feature is needed to make the segmentation system give robust results. To determine the thresholds we used the method in the study proposed by Giannakopoulos *et al* [**46**]*.* They used spectral centroid as the feature to represent the frequency characteristic of the data that will be segmented. After extracting the features, a dynamic thresholding method is applied. The threshold should change for every each audio sample used as an input to the segmentation system. In order to find the dynamic thresholds, the normalized histograms of the features are computed first. Then the local maximums are executed M_1 and M_2. Using weighted

averages of those maximums as in (3.13) we define the thresholds [**46**]. W is defined by user experimentally.

$$T = \frac{W * M_1 + M_2}{W + 1}$$

(3.13)

Then using this threshold for both features, voiced part decision is made. The red regions in Figure 3.8 are the voiced parts marked by our segmentation algorithm.
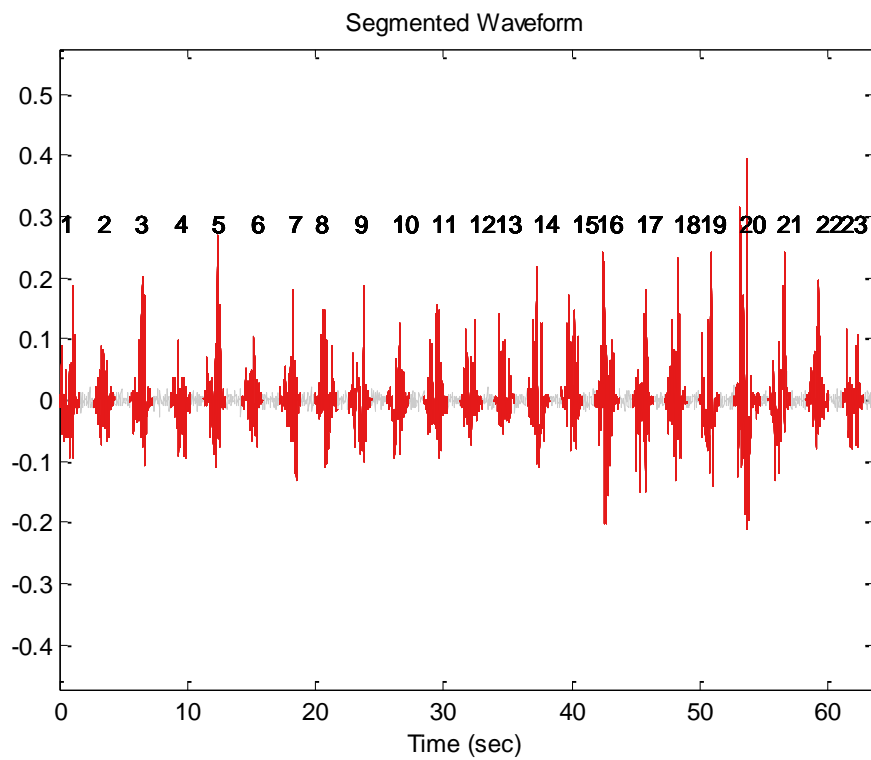
Figure 3.8: Segmented waveform

3.6.   Classification

Classification is dividing a given data set into *n* groups according to similarities of the values and/or features. For our problem, we aim to investigate the characteristics of different type of sounds, observe their similarities or differences by a feature based representation, and to detect special events like apnea or hypopnea by using these features. In order to achieve this goal, there is a strong need of a system that is able to show the behaviors of the different sleeping sounds and if possible catch those different sounds. This process can be done by supervised or unsupervised classification methods. In unsupervised methods, there is no or little prior knowledge about the input data, the grouping is done based on mathematical or algorithmic measures or similarities. On the contrary, in supervised case, there is a training set from which the classification algorithm attempts to infer a notion of similarity. Pre-defined classes are used as labels to label a data set called *training data set* and another data set called *testing data set* is classified according to the labeled training data set.

3.6.1. Unsupervised Classification Using K-Means

Unsupervised classification or clustering is the organization of patterns in the feature vectors into groups based on their similarities or differences. A cluster is a group of data that are similar to each other within the cluster and dissimilar to the data in other clusters. The input data for the clustering process is the frame based feature vector extracted from the sleeping sounds. Hierarchical clustering, self organizing maps (SOM), k-means and k-medoids are the well known and mostly used clustering techniques [**47**].

K-means is one of the most popular clustering algorithms because it has a simple implementation process and also it is computationally efficient [**48**]. This method first starts selecting randomly selected *k* data which are *d* dimensional (same as the dimension of each feature vector). Randomly chosen feature vectors are assigned to be the first cluster centers. Then the Euclidean distance is calculated to measure the distance between the data points and cluster centers. Each data object (feature vector) x_i is assigned to its nearest cluster center c_k. This separates the data in to k classes. After that, the algorithm updates each cluster centre c_k as the mean of all x_i that have been assigned to the corresponding cluster [**48**]:

$$D = \sum_{i=1}^{n} [\min_{k=(1..K)} d(x_i, c_k)]^2$$

(3.14)

$$c_k = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \qquad \text{for } 1 \leq i \leq k$$

where N_i is the number of data points in k_{th} cluster

(3.15)

This process continues until the clusters are stabilized. The square error information:

$$SquareError = \sum_{i=1}^{k} \sum_{j=1}^{N_i} |x_{ij} - c_k|^2$$

(3.16)

can be used as a criteria of how well the clustering is performed.

The feature set used in the classification algorithm contains many features that characterize different properties of sounds; the values of these feature vectors are in the order of 10^{-3} to 10^4. Therefore, higher values may dominate the distance and clustering would give wrong results. In order to avoid this situation, normalization is applied before applying the k-means algorithm. We convert the

features set in to a zero mean and unit variance feature o prevent the large values dominate the distance measure.

However k-means has some drawbacks, and various studies have been carried out to overcome these drawbacks. The initialization is a very important issue in k-means. As described in the algorithm, the number of clusters and initial cluster centers are needed to be defined by the user. This is mostly done randomly, but this affects the clustering result; in the iterative process the algorithm may stop at a local minimum which would give wrong clustering result. There are additional algorithms in the literature to overcome this problem [48]. Besides, random initialization would result different cluster labels for every different execution of the algorithm. This is not a big issue for our case as we just use this algorithm to group sounds and see the differences in their distribution. There are also algorithms to select an appropriate number of clusters in literature [48], however we did not consider those algorithms here since we used k-means clustering algorithm for comparison only; we used more advanced grouping methods for further steps which will be explained in the following part.

3.6.2. Supervised Classification Using Gaussian Mixture Model (GMM)

Supervised classification algorithms need ground truth data; these labeled data which are also represented by feature vectors, are used to train a classifier to label un known data. The success of the classification strongly depends on the quality of the training data. Support vector machines (SVM) [49], k nearest neighbor (k-NN) [50], neural networks [51], Gaussian Mixture Model (GMM) [50], Hidden Markova Model (HMM) [50] are the most popular supervised classification methods.

In audio classification GMM and HMM are very commonly used. The success of HMM and GMM are reported in improving the classification accuracy. This result

is clearly observed in Figure 3.9 where the error rate responses of various feature vector and classification methods to feature numbers is mentioned.
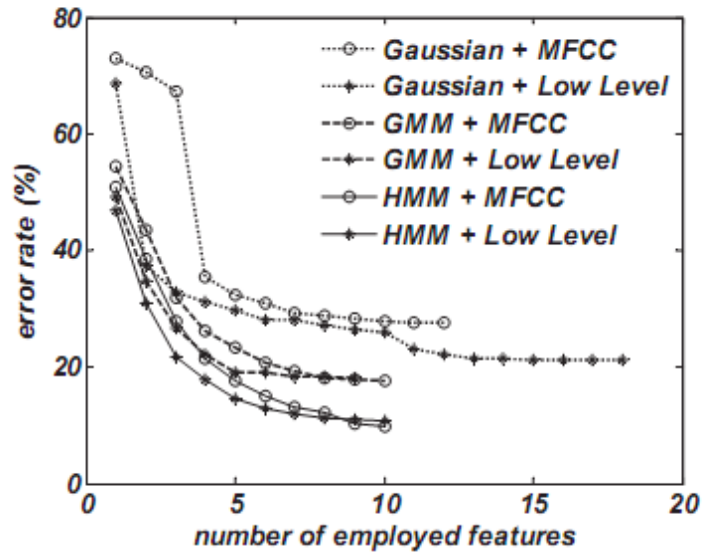
Figure 3.9: Error rate as a function of employed features [52]

Although HMM shows a better classification performance when compared to GMM as shown in Figure 3.10. It brings unwanted computational cost since it uses higher order feature vectors, which will definitely be more significant with the highly sampled snoring sound recordings. Also GMM does not require any Markovian constraints between the audio classes [**52**]. Thus, considering those factors, we continued our work with GMM.

.
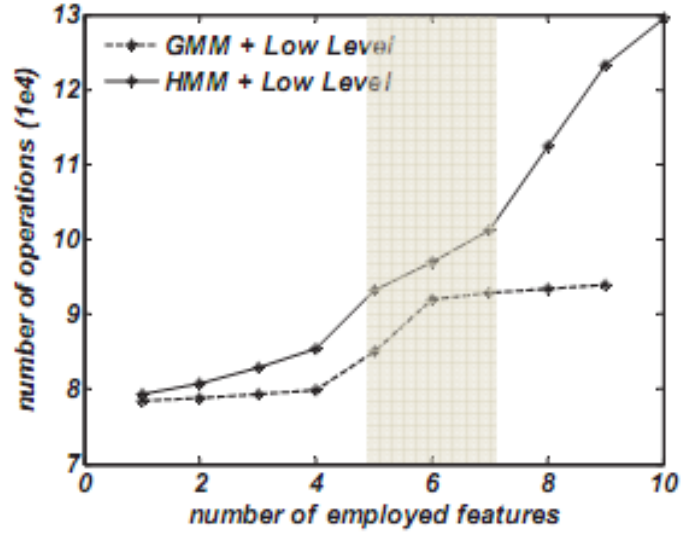
Figure 3.10: Computational cost as a function of employed features [52].

GMM is a classification algorithm that is based on modeling the pattern with a single or multiple Gaussian distributions [**50**]. Assuming that the feature vectors per each frame, $X = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots \bar{x}_m\}$ are statistically independent and identically distributed. The likelihood function related to each sleeping sound type is given in (equation).

$$p(X = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots \bar{x}_m\} \mid S_k) = \prod_{i=1,m} p(\bar{x}_i \mid S_k)$$

$$k=1,\ldots \text{number of sleeping sound classes} \tag{3.19}$$

We can assume that the likelihood of the vectors can be expressed with Gaussian distributions in *d* dimensional space (3.20).

$$p = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\{ \frac{1}{2}(\bar{x}_i - \mu)^T \Sigma^{-1} (\bar{x}_i - \mu) \} \tag{3.20}$$

$$\mu = \mathrm{E}[x_i] \tag{3.21}$$

$$\Sigma = \mathrm{E}[(x_i - \mu)^T (x_i - \mu)] \tag{3.22}$$

48

The Gaussian distribution can be represented with its mean (3.21) and covariance matrix (3.22). So the likelihood function is represented by those parameters [53].

In some problems only one single Gaussian may not be enough for representing the distribution. Therefore multiple Gaussians are constructed by linear combinations of n-Gaussian pdfs as in (3.23).

$$f_k(x) = \sum_{i=1}^{n} \pi_i p(x)_i$$

(3.23)

Where,

$$\sum_{i=1}^{n} \pi_i = 1 \quad \text{and} \quad \pi_j \geq 0 : j \in 1,.....n$$

(3.24)

π_j is the weighting component. It shows the importance of each Gaussian pdf. It either can be arranged as to make all to Gaussians equally weighted or it may be arranged differently [54]. In this case, π becomes another parameter for representing the likelihood function.
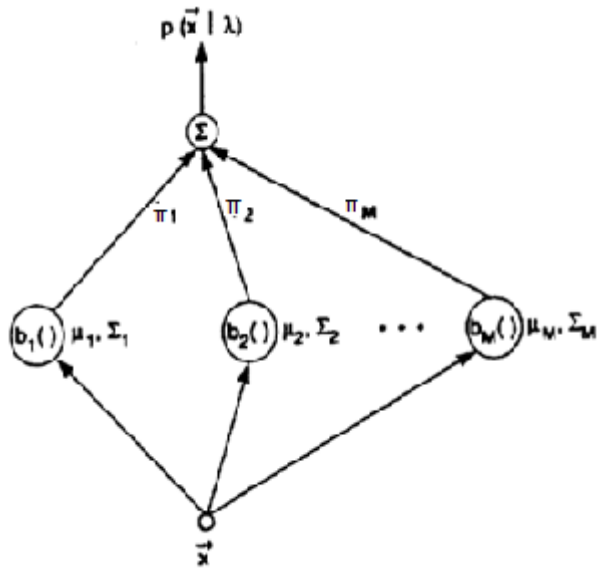
Figure 3.11: Scheme of GMM [54**].**

During training, we collect all the vectors for a given sleeping sound and our task is to learn the parameters of the Gaussian mixture, i.e. the mixing weights, the mean vectors and the covariance matrices. We achieve this goal using the well known Expectation-Maximization (EM) algorithm using expectation maximization (EM) algorithm [**55**]. In the E step, missing data are estimated given the observed data and current estimate of the parameters. In M step, the likelihood function is maximized assuming that the missing parameters are known from the E step [**56**]. Then the episodes represented by features are labeled by the class according to the maximum value it gets from.

3.7. Evaluation of the Results

To evaluate and comment on the classification results we need numerical representations. The statistical expressions which are defined for disease screening tests in [**57**] are used for this purpose. Before defining those expressions we need to define 4 parameters:

a: The number of successfully detected post-apnea episodes

b: The number of missed post-apnea episodes

c: The number of false alarms

d: The number of successfully identified non-post-apnea episodes

The statistical extracted from these parameters are listed below:

Sensitivity (Recall): It is defined as the probability that the test says a person has the post apnea and he/she has it. $\dfrac{a}{a+b}$

Specificity: It is the probability that the test says a person does not have post apnea episode and he/she does not actually have it. $\dfrac{d}{d+c}$

False Positive: It is defined as the probability that the test says there is a post-apnea episode but in fact it is not correct. $\dfrac{c}{c+a}$

False Negative: It is defined as the probability that the test says there is no post apnea episode but in fact there is. $\dfrac{b}{b+d}$

Also to measure the performance of the system, we can use *accuracy* measure which is the rate of correctly selected post apnea episodes among the ones that the test says it is a post apnea episode: $\dfrac{a}{a+c}$

From a successful system high sensitivity, specificity, and accuracy is expected where as the false negative and false positive values should be low.

Up to now the method is described in this chapter. In the following chapter we will observe the result of the whole algorithm and the sub parts separately.

CHAPTER 4

RESULTS

4.1. Flow of the Algorithm

Under the assumption that the post apnea sounds are detectable and differentiable from other snoring sounds. , we aimed to detect post apnea sounds and hence the apnea episodes using the algorithm given in Figure 4.1.
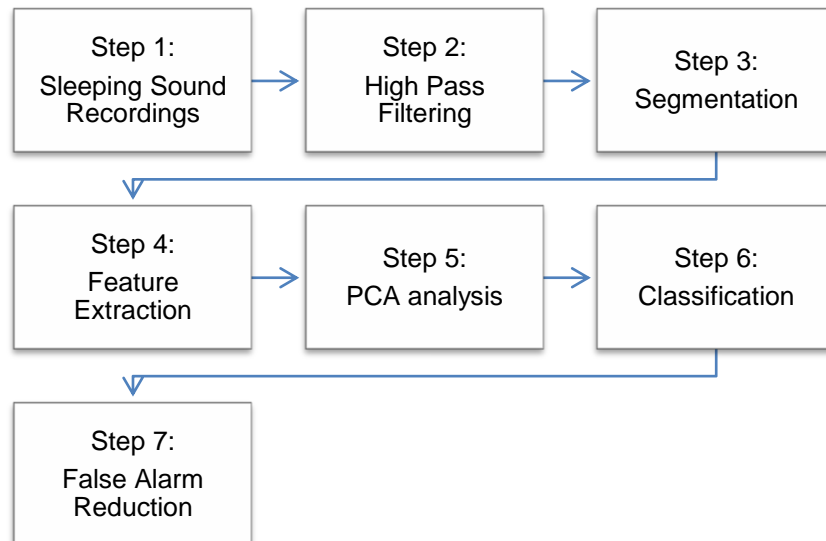
Figure 4.1: Flow of the algorithm

First the audio sample that is to be examined is selected. To eliminate the low frequency noise, high pass filter proposed in section 3.3 is used. The band edges of the FIR filter is determined using the information provided by the spectrogram plots of the recordings. As the recording environment is the same for all of the recordings, a single FIR filter is used for filtering all of the sleeping sound recordings. The stop band edge is defined as 60 Hz, the transition edge is defined between 60 Hz and 100 Hz, and the band from 100 Hz, to Nyquist frequency (8000 Hz) is defined as the pass-band. The characteristic of the filter is shown in Figure 4.2 The effect on an example audio sample can be observed from Figure 4.3. In Figure 4.3a, the time domain and the frequency domain characteristic of unfiltered waveform is seen. After applying the high-pass FIR filter, the new time domain and frequency domain characteristics are shown in Figure 4.3b.
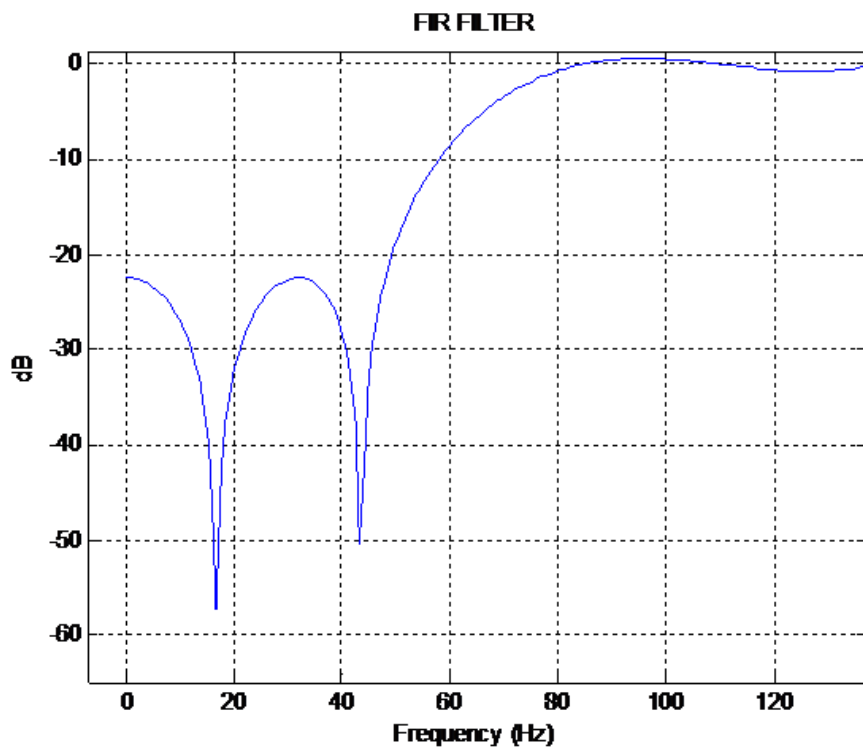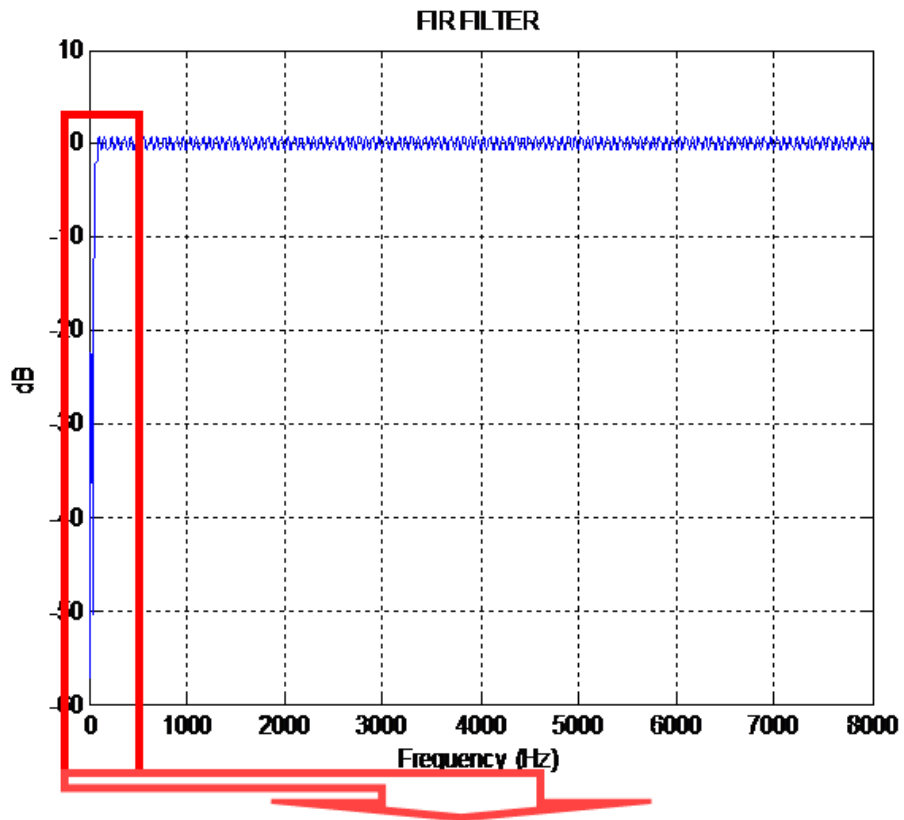
Figure 4.2: Frequency characteristic of FIR filter
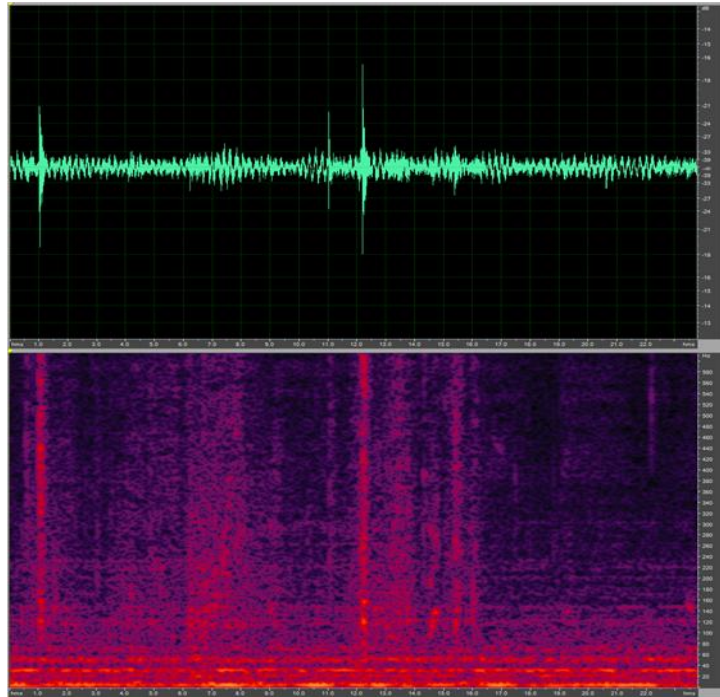
(a)
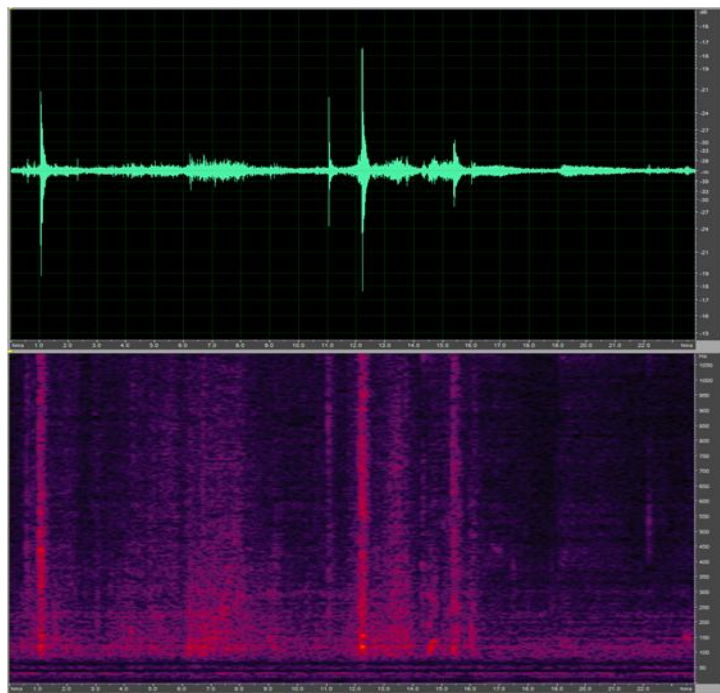
(b)

Figure 4.3: Time domain and Frequency domain characteristics of a waveform a) Before applying FIR filter, b) After applying FIR filter

4.2. Segmentation

The next step is the segmentation step. What we aim in segmentation step is to eliminate the silence parts of the sleeping sound recordings. Although selected features are able to differentiate the silence and voiced parts, applying classification algorithm to entire part of the selected recording would bring a lot of computational load. Besides, the beginning and the end points of the voiced parts are important for proper labeling of the post apnea and other snoring episodes. The proposed method in section 3.5 is applied on a random interval of data in Figure 4.4. The segmentation result is seen in Figure 4.5. The weighting factors used in equation (2.12) are determined experimentally. If we choose high weighting factors, more detail is included in the segmented parts which may include unwanted noise. Choice of low values as weighting factors may cause missing some episodes. All the selected episodes are labeled with a number to help in verifying the classification results, which will be explained in 4.4.
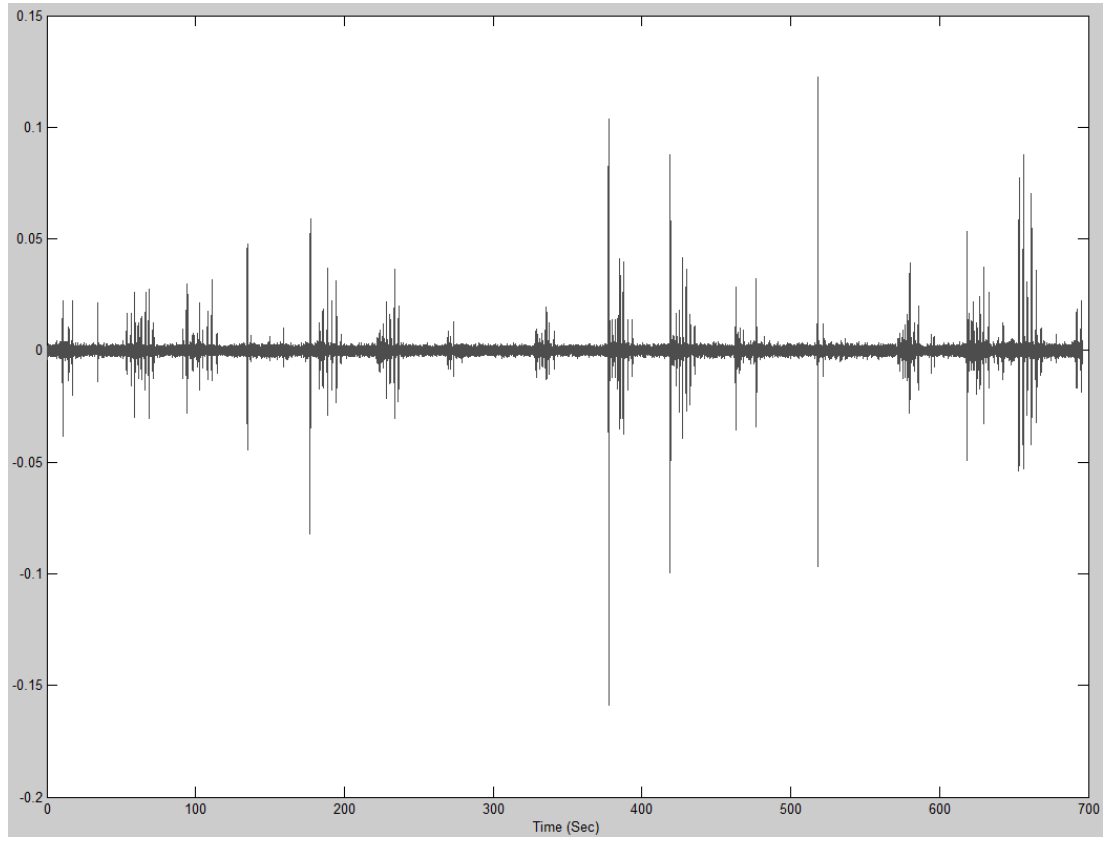
Figure 4.4: Waveform to be segmented
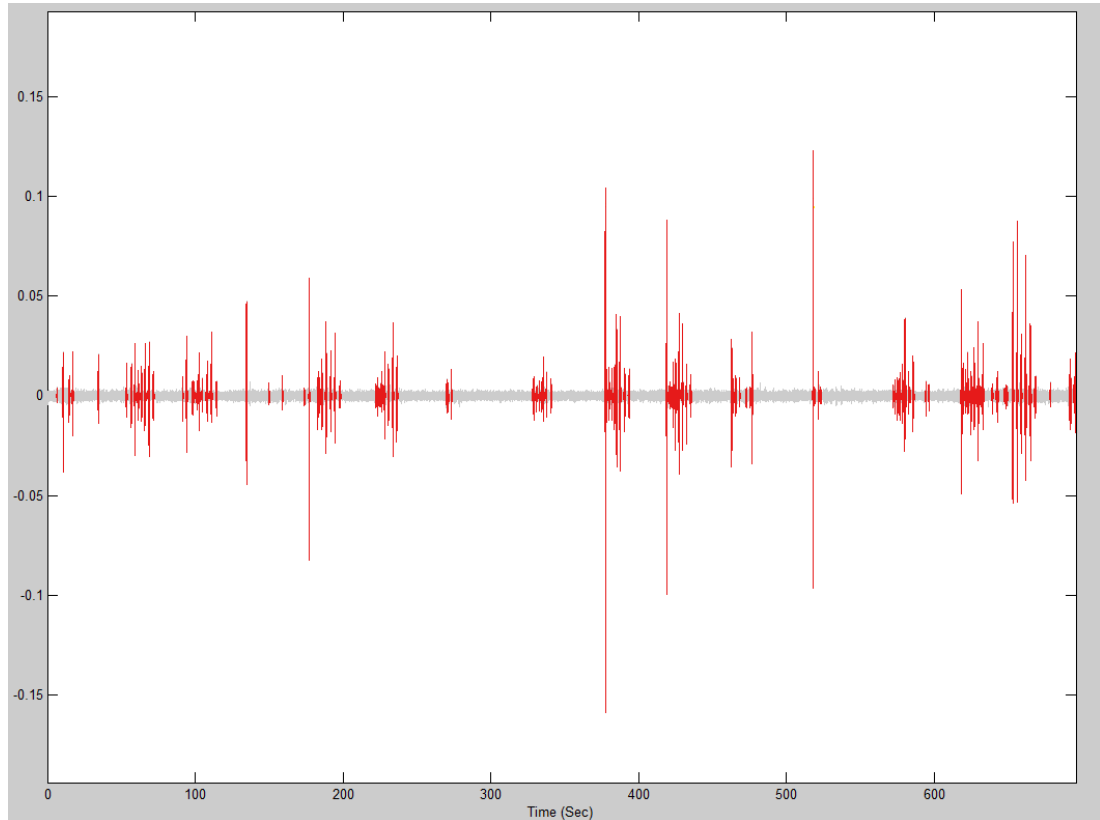
Figure 4.5: Result of the segmentation part

4.3.   Feature Extraction

After segmentation step, we obtain episodes with lengths of 0.5 sec to 3 sec. To classify those episodes, they need to be represented by proper features. The features introduced in Section 3.4 are extracted from 50 msec windows and a feature vector for each frame is obtained. Every episode contains more than one frame so the feature sets for each episode must be represented with some statistical properties of those features. Just using the mean or median of those features in each episode will cause loss of information as some characteristic properties of the sleeping sound episodes result from one or two special frames. Besides variation of the features give distinctive knowledge for the episodes.

59

While constructing the feature vectors for each episode, the acoustic differences of each class of sound must be examined very carefully. We are mainly considering normal snoring sounds, hypopnea period snoring sounds, post-hypopnea snoring sounds, apnea period obstruction sounds, post-apnea snoring sounds and a class of other noises consisting of human speech sounds and environmental sounds like door slamming sounds, bed sounds due to movements during sleep, sounds caused from objects accidentally falling down. Considering these facts, we examined maximum two, minimum two, mean, standard deviation and standard deviation over mean values of all the features described in Chapter 3 and constructed the feature vector which includes the mean of zero crossing rate, spectral centroid, spectral roll-off point, end and initial points of the LPC error curves, standard deviation of spectral flux, initial and end points, and the time constants of the LPC, mean of maximum two values of energy entropy error curve standard deviation over mean of short time energy.

The standard deviations of some of the features are selected because their variation gives more distinguishing information than their mean values. For the LPC parameters, standard deviation of those values enables the amount of harmonicity changes in the sleeping sound recordings. The amount of change in spectral change speed also gives valuable information as it will have lower values for the episodes whose spectra do not change a lot.

The standard deviation over mean of energy is selected because as we are using different recordings we would need normalization of the waveforms if only mean or standard deviation of the energy feature is used.

4.4.   PCA for the Feature Set

The size of the feature vector may be unnecessarily large, as some of the features may carry same information. Mathematically, PCA is for eliminating correlated values in the feature vector, and for decreasing computational load as mentioned in Section 3.4.9. The size of the feature vector is decreased to 10

with this analysis. While decreasing the size, we choose the threshold quite large for the choice of dimensions to be eliminated not to lose much information.

4.5. Detection of Post Apnea Snore Episodes

In this part first we will show that the post-apnea sounds are detectable and they are grouped in to subclasses with unsupervised methods. Then, with supervised classification method, GMM the performance of our algorithm in detecting apnea periods will be discussed.

4.5.1. Unsupervised Detection of Post Apnea Sounds

Post apnea snoring sounds tend to have different characteristic from other type snoring sounds. This characteristic difference make them detectable in terms of the proposed features in Section 4.3 which will allow catching the apnea periods.

The audio sample given below contains 13 apnea periods and 13 post-apnea snoring episodes. To show that those episodes are detectable, we first choose the appropriate data set as described in 4.3. The episodes are clustered with the unsupervised k-means algorithm in to 6 clusters. Post-apnea episodes are known beforehand from the annotations done in GMMA. These post apnea sounds are marked with red circles in Figure 4.6. As it can be seen in Figure 4.6, most of the post-apnea episodes are grouped in a different cluster from the others.
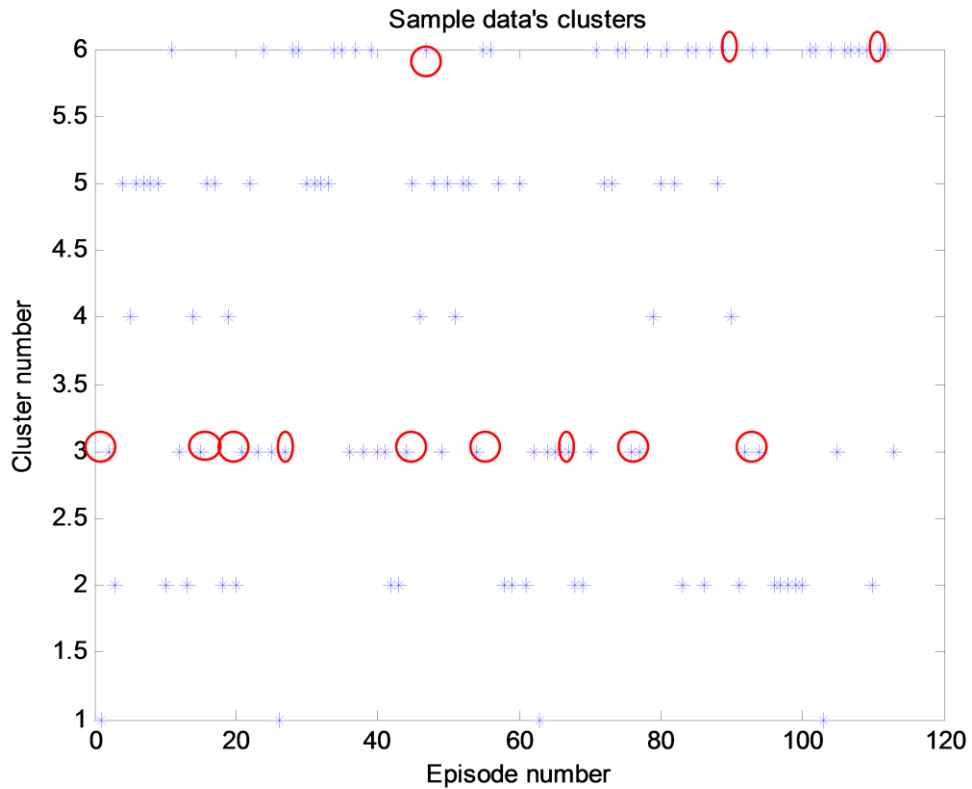
Figure 4.6: K-means clustering results of the sleeping sound sample

As seen in Figure 4.6 all of the post-apnea episodes are not grouped under one cluster as their characteristic may vary due to obstruction type of that moment. This variation can also be observed from Figure 4.7. In that figure, only the known post - apnea sounds in Figure 4.6 are clustered. As a result, the post apnea clusters of the audio sample do not compose only one single cluster. This result shows that post-apnea clusters carry different characteristics from other sleeping sounds in terms of the feature vector representation in Section 4.3. Also from that result it can be seen that some post apnea sounds may be similar to other type snoring sounds, which may cause classification errors in apnea detection.

One other result that can be obtained from Figure 4.6 is that hypopnea period sounds and normal snoring sounds are not separable in terms of our representation. This supports the expectation of the medical doctors in GMMA that they are inseparable by investigating sound recordings. They are mostly

grouped under 2nd and 3rd clusters, but no clear separation between them is observed.
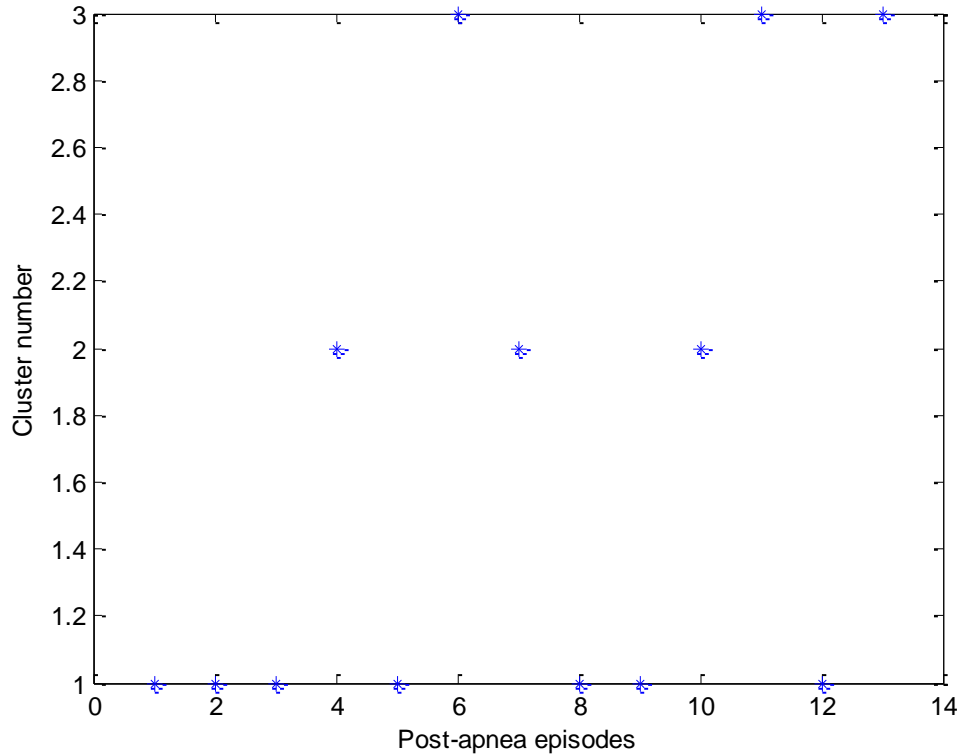
Figure 4.7: Post apnea sounds carrying different characteristics

4.5.2. Supervised Detection of Post Apnea Sounds and Apnea Periods

In unsupervised clustering we have seen that post-apnea sounds carry different characteristics. But without a prior knowledge it is not possible to say anything about which class does the episode belongs to. More than knowing the selective structure of post-apnea sounds, we have features to distinguish non-human sounds, and non-snore sounds up to a level of success. Considering these, we aim to label the sample data in terms of normal snore (n), post-apnea snore (a),

intra apnea voices (x), hypopnea period snoring sounds (h) and post-hypopnea sounds (c) and other sounds (j).

After feature extraction the training data set defined in Figure 4.8 is used to label the sample data. The episodes in the training data are selected from the same patient's recordings but the episodes in the sample test data are not included in the training set. The training data set contains 15 episodes for each class, which makes a total of 90 episodes.
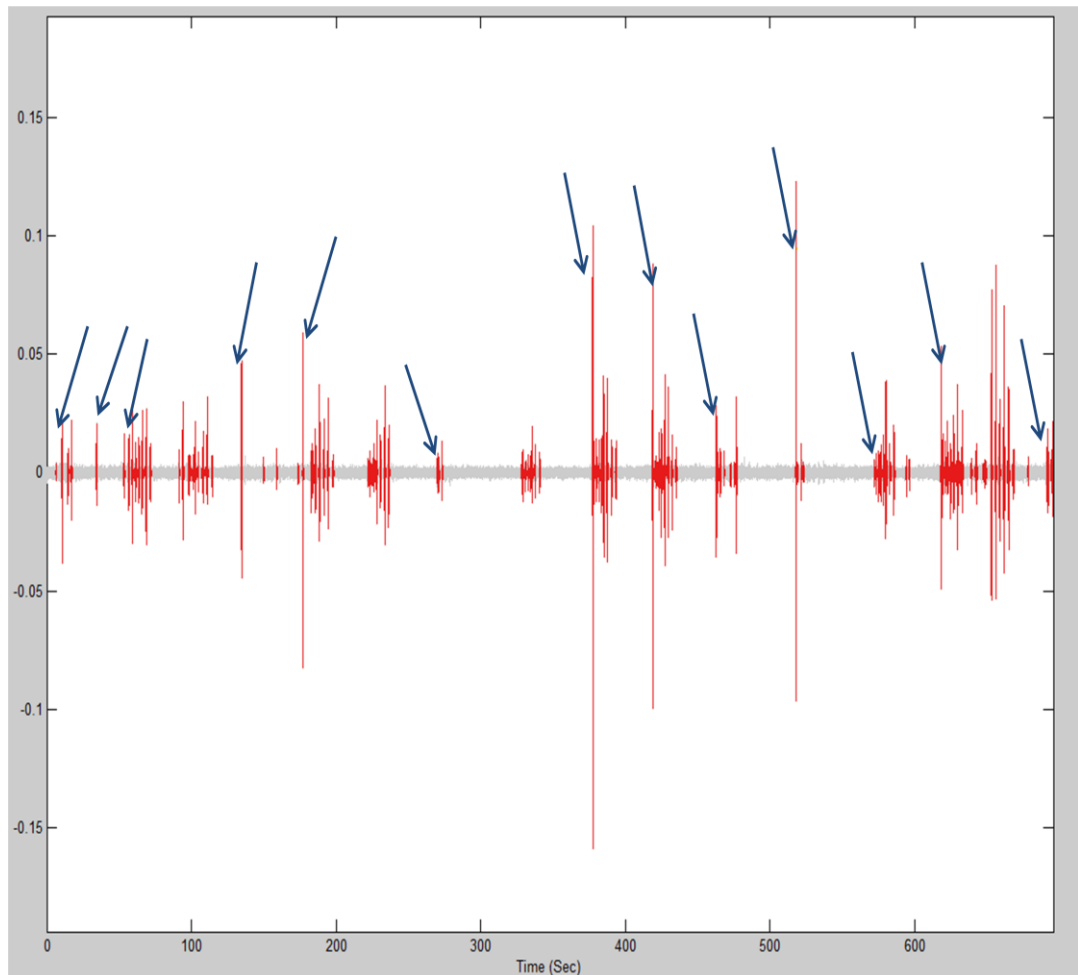
Figure 4.8:Sample data and some of the post apnea sounds in that data

Table 4.1: Labeling results of the sample data

EPISODES	LABELS									
1-10	'x'	'a'	'n'	'h'	'a'	'h'	'h'	'c'	'a'	'n'
11-20	'n'	'n'	'n'	'a'	'a'	'h'	'x'	'n'	'a'	'n'
21-30	'h'	'h'	'a'	'n'	'x'	'j'	'a'	'n'	'c'	'h'
31-40	'n'	'h'	'h'	'x'	'x'	'x'	'x'	'x'	'x'	'a'
41-50	'a'	'n'	'n'	'a'	'a'	'a'	'x'	'a'	'x'	'h'
51-60	'h'	'a'	'a'	'a'	'c'	'c'	'h'	'n'	'n'	'n'
61-70	'n'	'a'	'c'	'n'	'a'	'c'	'c'	'n'	'n'	'a'
71-80	'c'	'h'	'x'	'x'	'n'	'h'	'a'	'c'	'a'	'x'
81-90	'a'	'n'	'n'	'n'	'h'	'n'	'n'	'n'	'a'	'x'
91-100	'x'	'a'	'n'	'x'	'c'	'n'	'n'	'n'	'j'	'n'
101-110	'n'	'n'	'n'	'n'	'a'	'c'	'c'	'n'	'n'	'j'
111-113	'a'	'n'	'x'							

Table 4.2: (a): Test results for the supervised classification of sample data (without applying false alarm reduction algorithm) (b): Statistical test results for the supervised classification of sample data (without applying false alarm reduction algorithm)

		TEST RESULTS	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	13	0
Status	**No Apnea (-)**	11	89

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
1	0,542	0,890	0,458	0

(b)

From

Table 4.2, it is observed that the accuracy rate is low and the false positive rate is high due high number of post-apnea episodes that are not actually post apnea episodes which we name as "false alarms". They may carry similar characteristics but they are not actually the first sound episodes after the apnea-period. After apnea period ends, the patient may not immediately start breathing or snoring normally. He can continue making abrupt noises, so some of the false alarms are because of this reason. Besides, the feature vectors may not distinguish some of the external noises and apnea-period noises with post apnea episodes due to lack of information about those classes. Also, patient sometimes may snore or make similar noises like the post-apnea episodes or some of the post apnea episodes may be like other snore episodes due to unknown reasons, which causes a very small ratio of false alarms.

In order to decrease the ratio of those false alarms, we propose to detect the length of the silence part or the obstruction sounds possibly occurring during the apnea period before those episodes that are assumed to be post- apnea episodes. We know that apnea is defined as the period of blockage in breathing about 10 sec. or more and a related silence period. So defining a threshold for the length of the silence part before the post-apnea labeled episodes may help in correct decision of where the apnea occurs. Also in some situations patients may make obstructive sounds. So, detection of any obstruction sound may decrease the rate of false alarms. Using above two constraints the performance of the apnea detection algorithm may be improved. The result of the false alarm reduction algorithm is in Table 4.3. The accuracy rate increased and the false positive rate is decreased. The drawback of this method is that it can increase

the false negative ratio and decrease sensitivity due to increased number of missed post-apnea episodes. The misses are mostly because of lack of information about the apnea period sounds.

		TEST RESULTS	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	12	1
Status	**No Apnea (-)**	3	81

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,923	0,800	0,970	0,200	0,010

(b)

Table 4.3: (a): Test results for the supervised classification of sample data (after applying false alarm reduction algorithm) (b): Statistical test results for the supervised classification of sample data (after applying false alarm reduction algorithm)

4.6. Experiments

In order to determine the success of our algorithm we applied a series of tests using different training data sets. The testing data set consists of two different audio parts with a duration of 15 min, chosen from two different patients. While choosing those data sets we aimed to include as much apnea periods as possible and tried to keep the variety of sleeping sound as much as possible.

We applied 5 experiments for 5 different training sets for each testing data.

CASE - 1: For the first case, the training data set consists of the sleeping sound episodes randomly taken from the tested patient's recordings itself. While

choosing episodes for the training set, the episodes of the recording used as the test data are not included. Training data set for Patient -1 contains more post - apnea data that the training set for Patient - 2 as many of the annotated post apnea sound is selected in the test data.

CASE - 2: The second training set includes data that have clearly detectable post apnea snoring episodes. Also those episodes carry similar characteristic with the testing data when listened to. The training set for this case contains episodes taken from two different patients other than the tested patients.

CASE - 3: For the third case, the experiments are carried with the training sets that contain episodes from all the available data in hand including the tested patient's snoring episodes. However, the episodes of the tested audio parts are excluded from the training sets. Same training sets are used for this case. The set contains maximum number of data, 360 samples total.

CASE - 4: The fourth case presents a more realistic scenario. Again, the experiments are carried out with the training sets that contain all the available patients' snoring episodes, but this time, the data from the patient that is being used for testing are excluded.

CASE - 5: Some of the recordings are very noisy or the episodes can't be clearly observed due to recording *faults. But in real life those situations may occur, so the fifth case is to observe the algorithm's reaction against such conditions.* The training sets for both cases is the same.

4.6.1. Test Results

CASE - 1:

Table 4.4: (a): Test results for the supervised classification of the sample taken from patient-1 for case-1(b): Statistical test results for the supervised classification of the sample taken from patient-1 for case-1

		PATIENT - 1	
		Positive (+)	**Negative (-)**
True	**Apnea (+)**	13	1

Status	No Apnea (-)	6	249

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,929	0,684	0,977	0,326	0,004

(b)

Table 4.5: (a): Test results for the supervised classification of the sample taken from patient-2 for case-1(b): Statistical test results for the supervised classification of the sample taken from patient-2 for case-1

		PATIENT - 2	
		Positive (+)	**Negative (-)**
True	**Apnea (+)**	24	7
Status	**No Apnea (-)**	4	174

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,774	0,857	0,978	0,143	0,039

(b)

The training sets are constructed by the rest of the both same patient's recordings for each patient. The specificity rates are similar as seen from Table 4.4 and Table 4.5. In patient-1's results the false positive value is more than the patient-2's. This is because patient-1's recording is chosen from a part that the variety of sounds is very high when compared to patient-2's recording. On the other hand the accuracy rate of patient-2 is higher than the accuracy rate in patient-1's. Because the tested part of patient-2 is chosen from a part that includes many post apnea sounds the rest of the post apnea sounds that is out in the training data set may not be sufficient. In patient-1's recording the training

set contains more post-apnea sounds than the patient-2's first training data set. So the difference in accuracy rate is an expected result. Due to large number of missed post apnea episodes, the false negative ratio for patient-2 is larger than patient-1's.

CASE - 2:

Table 4.6: (a): Test results for the supervised classification of the sample taken from patient-1 for case-2(b): Statistical test results for the supervised classification of the sample taken from patient-1 for case-2

		PATIENT - 1	
		Positive (+)	**Negative (-)**
True	**Apnea (+)**	11	3
Status	**No Apnea (-)**	4	251

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,786	0,733	0,984	0,277	0,012

(b)

Table 4.7: (a): Test results for the supervised classification of the sample taken from patient-2 for case-2(b): Statistical test results for the supervised classification of the sample taken from patient-2 for case-2

		PATIENT - 2	
		Positive (+)	**Negative (-)**
True	**Apnea (+)**	26	5
Status	**No Apnea (-)**	7	171

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,839	0,788	0,961	0,212	0,028

(b)

The patient-1sensitivity is decreased because of the number of missed episodes increased slightly. This is because the training set consists of two different patient's recordings. For patient-2 the change in the results is different from patinet-1.The accuracy rate of patient-2 decreased in experiment-2. This because the training data set contains episodes from other patients. But on the other hand the number of missed episodes decreased as there are more post-apnea sounds in the training data set, so the training is done better.

CASE - 3:

Table 4.8: (a): Test results for the supervised classification of the sample taken from patient-1 for case-3 (b): Statistical test results for the supervised classification of the sample taken from patient-1 for case-3

		PATIENT - 1	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	13	1
Status	**No Apnea (-)**	7	248

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,929	0,650	0,973	0,350	0,004

(b)

Table 4.9: (a): Test results for the supervised classification of the sample taken from patient-2 for case-3 (b): Statistical test results for the supervised classification of the sample taken from patient-2 for case-3

		PATIENT - 2	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	25	6
Status	**No Apnea (-)**	2	177

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,807	0,962	0,994	0,039	0,033

(b)

Number of false alarms increased for patient 1, which can be observed by decrease in accuracy and specificity. This is a normal result because the variety of episodes is increased in this training data set. On the other hand the accuracy increased and the false positive rate decreased for patient 2. The decrease in false positive is an expected as it is normal to have less missed episodes. That much increase in accuracy can be explained by having similar external noise or similar apnea period sounds as this is a result of false alarm reduction algorithm.

CASE - 4:

Table 4.10: (a): Test results for the supervised classification of the sample taken from patient-1 for case-4 (b): Statistical test results for the supervised classification of the sample taken from patient-1 for case-4

		PATIENT - 1	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	11	3
Status	**No Apnea (-)**	5	250

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,786	0,691	0,981	0,313	0,012

(b)

Table 4.11: (a): Test results for the supervised classification of the sample taken from patient-2 for case-4 (b): Statistical test results for the supervised classification of the sample taken from patient-2 for case-4

		PATIENT - 2	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	25	6
Status	**No Apnea (-)**	5	173

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,809	0,833	0,972	0,168	0,034

(b)

This case is the most similar one to the practical case as it does not contain any episode from the tested data and the variety of episodes for each class is very high. The false alarm rate increase in patient-2 which appears as the decrease in accuracy compared to the third case. Also observing that the accuracy increased for patient-1 which is an expected due to same reasons in case-1 and case-2. The decrease in accuracy for patient-2 shows that the apnea period sounds or the external noises are well defined that they decreased the false alarm rate in case 3 and case 1 compared to case 4.

CASE - 5:

Table 4.12: (a): Test results for the supervised classification of the sample taken from patient-1 for case-5 (b): Statistical test results for the supervised classification of the sample taken from patient-1 for case-5

		PATIENT - 1	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	12	2
Status	**No Apnea (-)**	11	244

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,857	0,522	0,961	0,478	0,008

(b)

Table 4.13: (a): Test results for the supervised classification of the sample taken from patient-1 for case-5 (b): Statistical test results for the supervised classification of the sample taken from patient-1 for case-5

		PATIENT - 2	
		Positive (+)	**Negative (-)**
True	**Apne (+)**	24	7
Status	**No Apnea (-)**	15	163

(a)

Sensitivity	Accuracy	Specificity	False Positive	False Negative
0,774	0,615	0,926	0,338	0,041

(b)

For both patients the accuracy rate decreased a lot. When the recordings that constitute the training data set are listened it can be observed that they are not as clear as the in the other recordings. So this lack of quality in the recordings effected the result. The most significant impact of this case is that the false alarm rate increased which can be detected from the decrease in accuracy.

To observe the changes of statistical values in between cases and the two patients reported previously are visualized in the histogram plots in Figure 4.9 to Figure 4.13.
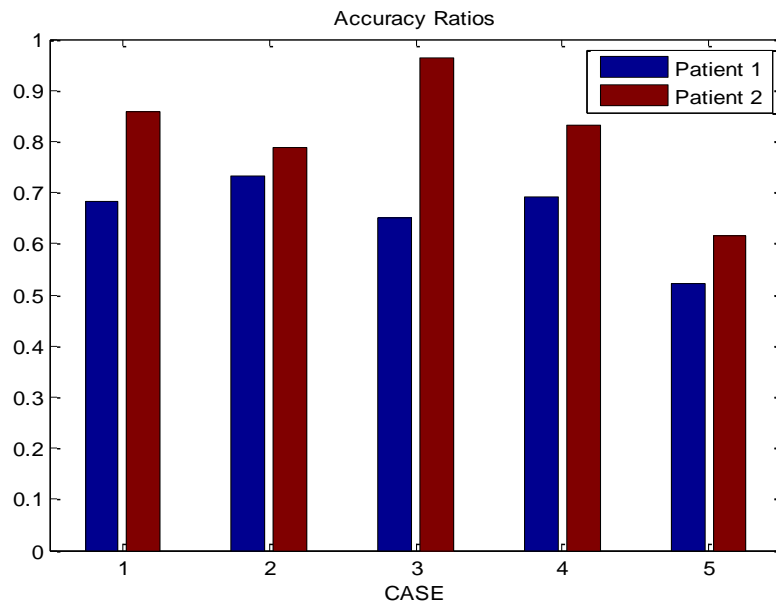
Figure 4.9: Accuracy ratios of the cases applied on two different recordings
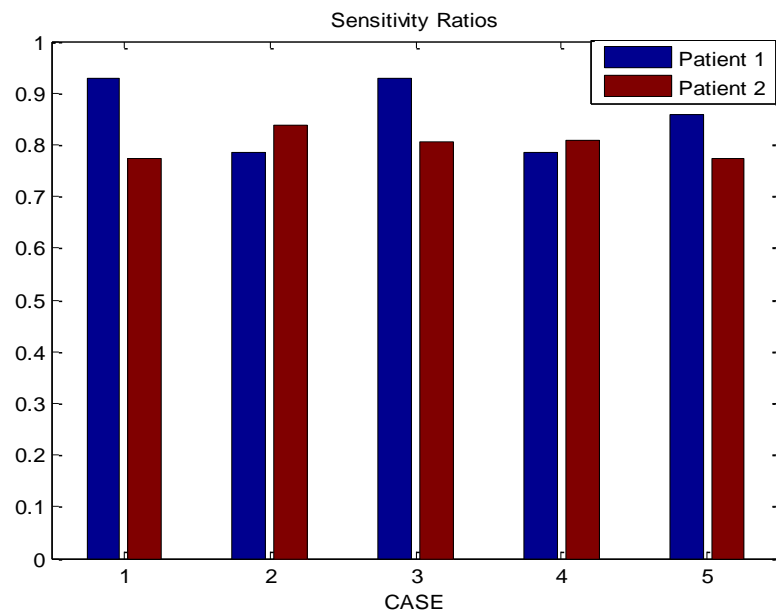
Figure 4.10: Sensitivity ratios of the cases applied on two different recordings

75
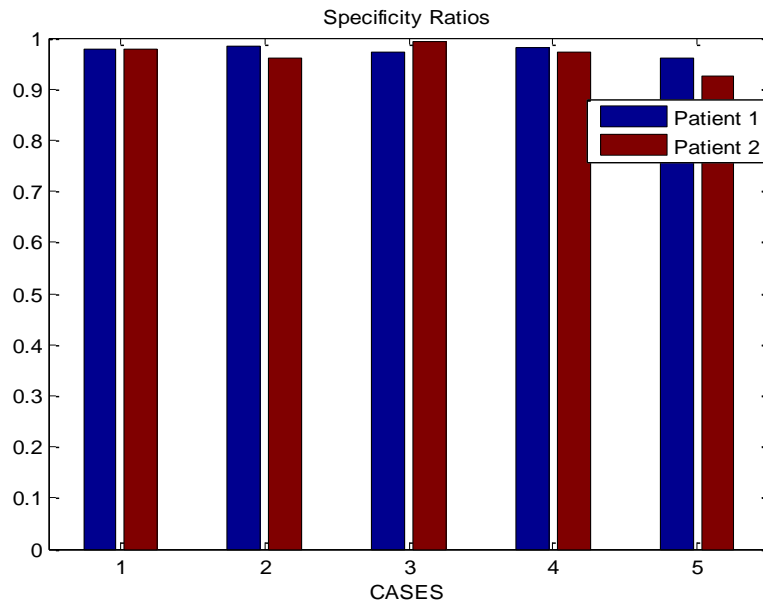
Figure 4.11: Specificity ratios of the cases applied on two different recordings
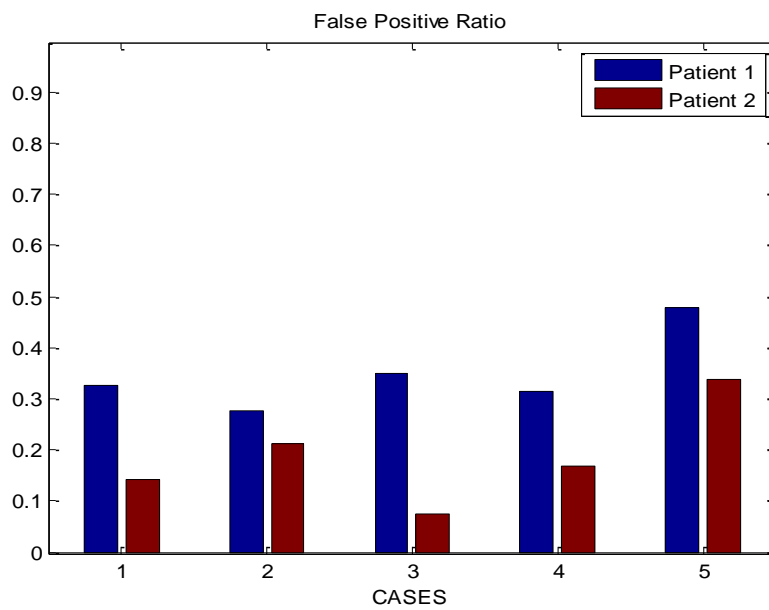
Figure 4.12: False positive ratios of the cases applied on two different recordings
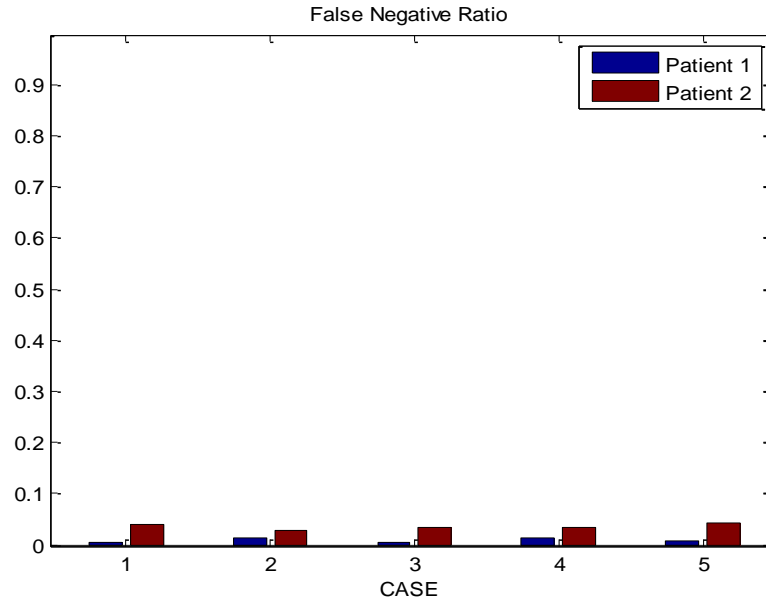
Figure 4.13: False negative ratios of the cases applied on two different recordings

4.7.   Discussion

We have done several experiments for different cases. The cases are arranged with different difficulty levels in order to observe several different results.

The performance of the algorithm examined by getting the statistical ratios described in Section 3.7. We expect high accuracy, sensitivity, specificity and low false positive and false negative values from the results to say that the algorithm succeeded.

At overall view, the experiments done with patient-2 seems more successful. This is because there are many post apnea episodes in the testing data set. So any change in number of truly detected post apnea episodes, missed post apnea episodes or false alarms change the statistical properties a lot. While looking at the results this fact should be observed.

Also, the detected number of post apnea didn't change a lot in different cases, which shows the stability of the algorithm. This can be observed from the almost stable and high specificity numbers. On the other hand the variation is higher in false alarm. This is because lack of information for the apnea period sounds and external noise as existence of them is very important for false alarm detection.

The most realistic case is *case-4* as the training data set contains data sets other than the patients episodes and also contains many episodes with many different kind of post apnea episodes.

There are many missed post apnea episodes in *case1-patient2.*This is due to the fact that, we choose many post-apnea episodes of that patient to be tested and less remained to be used in training data. Considering the variety of post apnea sounds for a single patients recordings taken in one night and need of lots of training data for supervised classification problems, the predicted success couldn't be obtained from that experiment. In c*ase4-patient2* some statistics are better from *case1-patient2* which shows the number of episodes in training data set and the variety of the episodes in the training data set is more important than having similar data in both training and testing data sets.

While getting conclusion from these results, we need to consider that increasing number of episodes in a data set would increase the accuracy of the classification results. Obviously the data used in training of the classifier is not enough especially for some cases, so this causes decrease in performance.

Also, the segmentation result would affect the accuracy of the results. With a more detailed segmentation some meaningless audio parts may be chosen as episodes. So false labeling of those episodes may cause false alarms or missed apneas due to false alarm reduction algorithm.

In some cases, especially for patient 1, quite low accuracy values is a fact to be considered. After the apnea period we only consider the first post apnea episode in classification. But in most cases this is not an only single episode. The patient continue to voice some noises in the following seconds of the first post apnea sound. Those sounds some times are labeled as in apnea and post apnea sounds which increase the rate of false alarms. There should be a class

definition for those sounds so that the performance of the detection algorithm can be improved.

During apnea period, normally a total silence is expected by the medical doctors. But due to obstruction some voices like chocking or whistles occur sometimes while the patient tries to take breath. The definition of those sounds should be done very carefully as occurrence of those sounds is one of the basic constraints in our false alarm reduction algorithm. Unfortunately, there is a lack of information for those sounds in the literature. Also, medical doctors could not guide us a lot about those sounds. Despite these, we couldn't exclude those apnea period sounds from our training sets in order not to miss apnea periods. Additional information should be gathered about those sound, so that acoustic features can be selected in order to differentiate those sounds also and more accurate solutions can be obtained.

One other problem is that there are lots of external sounds in the recordings. Also the variety of those sounds is more than expected compared to an average sleeping environment. The recording room was not in an isolated place and also the room was not sealed well against that much noise. Those external sounds caused miss labeling and false alarms a lot in our tests. We tried to include some of those noises in the training set and proper audio feature are tried be used to distinguish them but high variety and high quantity of those caused miss classification results, especially false alarms. Also like the apnea period sounds not using them in the training set missed episodes which is also an unwanted result. In our study we tried to use everything as in real life as possible so we didn't ignore those voices.


4.8.   Decision Guide of Unknown Episodes

For some situations, PSG recordings may be inconsistent. For example, the flow meter may be showing some activity which indicates that respiration is

continuing but the thermo sensor may detect no heat change to confirm that activity. Or the level of oxygen level decrease may not be helpful in saying that the period is apnea or hypopnea. In such cases, additional information is needed to make a decision.

Our algorithm is able to label the sleeping sounds for a degree of success. The labels are assigned by using the probability results obtained from GMM. Each maximum likelihood function defined for each class, described by (3.18) gives the probability values of belonging to that class, for every episode (represented by a feature vector) . In Figure 4.14, we can see an application of this property for one of the patients in our data base. While annotating the sleeping sounds of that patient, medical doctors could not decide on the class of some sounds. The probabilities for each of the five candidate classes are visualized by normalized histogram plots in Figure 4.14. First four sound episodes are most probably post-hypopnea sounds, and the last one is found as an obstruction sound occurring during the apnea period. Although it was annotated as a post apnea sound, the fourth episode was noted to be similar to a normal snoring or hypopnea period rather than being a post apnea or post hypopnea sound snoring sound by the medical doctors. This fact is supported by the histogram corresponding to this episode in Figure 4.14 where second highest probability value belongs to hypopnea period sound.
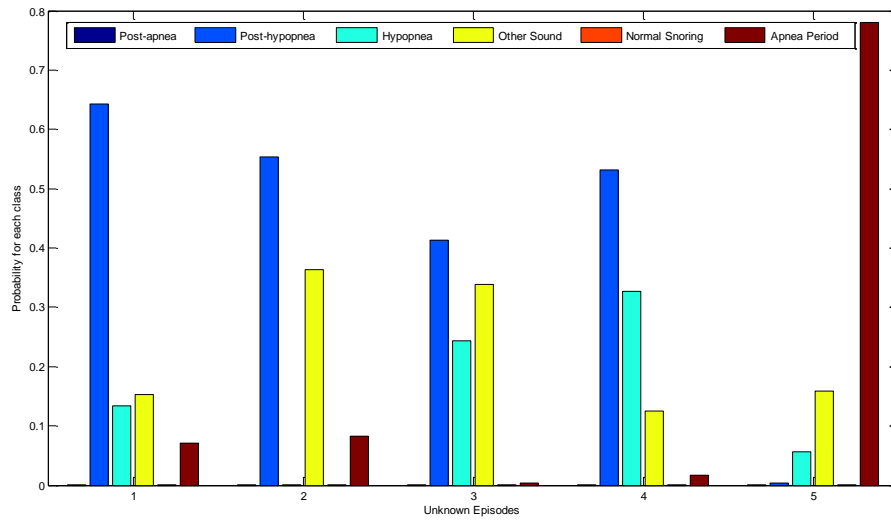
Figure 4.14: Probability values for unknown episodes

In the next example in Figure 4.15, a set of probabilities of six unknown episodes are shown. These are also reported as either post-hypopnea or post-apnea but can't be decided implicitly. So using the probability values extracted from GMM we are able to give a decision aid as in Figure 4.15. The sixth episodes is most probably a post-apnea sound and the rest are most probably a post-hypopnea sound as the result of our classification system.
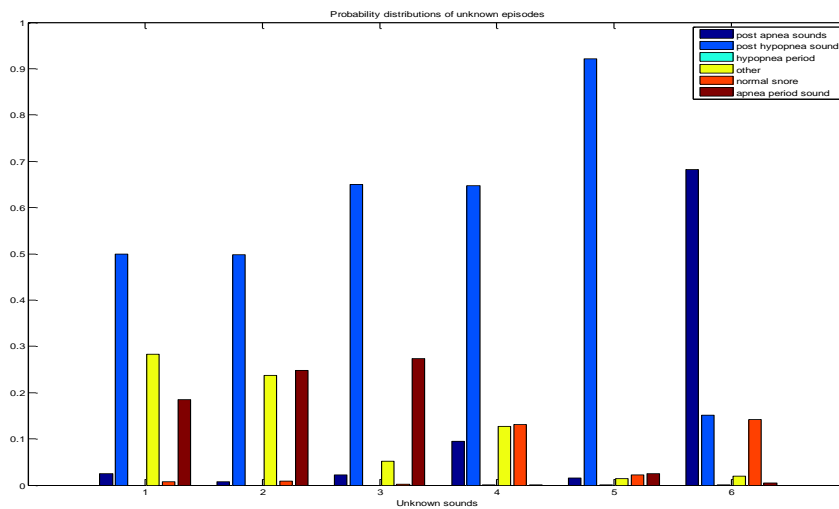
Figure 4.15:Probability values for unknown episodes

This may be a useful tool for the medical doctors if they come across confusing situations in making decisions upon examining PSG recordings.

CHAPTER 5

CONCLUSION

Sleep apnea is a disease that appears as unexpected and quite long breathing pauses during sleep. Those pauses have negative effects on patient's social life and cause serious health problems, it can even end with death. Treatment of apnea is possible but first it should be diagnosed. The diagnosis step of this disease is somehow difficult, because it is an anomaly that takes place during sleep. The drawbacks of classical diagnosis methods are mentioned before. In recent years, the researchers are examining snoring sounds in order to create alternative diagnosis method for apnea. In this thesis, we aimed to make contributions in diagnosis of sleep apnea by investigating sleeping sound recordings taken with non-contact microphones.

Our starting point was to investigate how post-apnea sounds can be used for diagnosis of apnea. To achieve this goal, first we need to show that the post apnea sounds have different characteristic from other sleeping sounds that make them detectable. For this purpose appropriate features are adopted from the past studies in the same area and also in speech recognition, various audio segmentation studies. In addition to these known features, we implemented new features, which can be used to differentiate the sleeping sounds into different classes. To show that the selected features work for discriminating sleeping sounds, also proper classifiers are needed. Extraction of features from the sleeping sound recordings and implementing classification algorithms that use those features forms the main part of our study. Also, pre-processing for the recordings to cancel the low frequency noise, segmentation to obtain each sound episode separately and PCA analysis to decrease the dimension of the feature vectors constitute side parts of our algorithm.

The recordings taken from GMMA are carefully annotated with the guidance of medical doctors and we created our database. The database was somehow problematic as it was very noisy and hard to synchronize with PSG recordings.

After creating our algorithm and applying it to our database, we observed that most post apnea sounds were detectable with our system. The harmonicity of post apnea sounds are less and the center frequencies are higher than the normal snoring sounds and the hypopnea period sounds. Also energy variation in post-apnea sounds was higher than other kind of episodes which makes energy entropy and standard deviation of energy valuable features for our goal. Unfortunately, these properties of post-apnea sounds do not result in a single class that all the post apnea sounds can be grouped in. We observed that the post apnea sounds in a single patient's sleeping sound recordings can have different characteristics even in a short period of time. This may be because of many factors like sleep position, sleep stage, head position, conditions of muscles and tissues in the obstructed area, etc. Because of these effects classification may be more difficult.

With supervised classification method, the episodes in the sleeping sound recording can be labeled as post-apnea, normal snoring, hypopnea period snoring sound, post-hypopnea sound, apnea period sound and other sounds. Our features were selected in order to discriminate post-apnea sounds, but in a whole night sleeping sound recording, there are various sounds taken from a patient. Besides, environmental noises due to human activity near the recording room and human speech was unexpectedly high. Considering these facts, we improved our feature set with the features that are capable of representing the difference between those external noises and sleeping sounds. Although we obtained success with this change, the variety of noise was too much to be represented by our feature set. Also, including features for discrimination of external noises would decrease the performance of our system in discriminating post-apnea sounds and other sleeping sounds. Zero-crossing rate and spectral entropy were valuable features for this job, along with spectral roll-off point and spectral centroid features. So the feature set is constructed considering these situations.

The post-hypopnea sounds are quite separable in terms of our feature vector representation. There is no claim in detecting post-hypopnea sounds and the detection of those sounds was not as successful as in the post-apnea case but, as there is a partial obstruction during hypopnea, the characteristic of those hypopnea sounds are similar to the characteristics of post-apnea sounds. Furthermore, when we investigate hypopnea period sounds, we see that they are not separable from normal snoring sounds. This is an expected result, because partial obstruction of air-way is the reason of both situations.

During apnea, normally a silence period is expected, but in practice this is not the real case. As a normal reflex, patients try to breath during that period. With that attempt, some obstruction noises like choking can be heard. The definition of those noises is not clearly made and sometimes even medical doctors could not decide whether the period is hypopnea or apnea when they came across such noises. Although we tried to choose those sounds carefully for our training set, the lack of information about those sounds cause some errors in our results. A more detailed study should be carried out in order to reduce these errors.

While using GMM, there is an issue that needs attention. The decision for the choice of number of Gaussians to represent each class is a very crucial point. Representing the sound classes with a single Gaussian distribution was not sufficient, so we used multiple Gaussians. The number of Gaussians changed our results slightly, so we tried use optimum numbers for each training set. In each experiment as the variety of episodes in any class changes, the number of Gaussians needed to be adapted.

Due to previously stated reasons, our algorithm may miss some episodes and also gives false alarms in the results. Number of missed episodes is very small when compared to false alarms. The major reason for false alarms is that, when the apnea period ends, the patient cannot immediately settle back to its normal respiration cycle, and some post-apnea like sounds continue to be heard. Also the previously stated external noise problem was an important reason for those false alarms. We put some constraints for reducing those false alarms. The major constraint is the detection of silence length before the post apnea episode. Unfortunately we cannot talk about an absolute silence for all of the apnea

periods so detection of apnea period sounds before the post apnea sounds is taken as a successful result also. This method worked well and decreased the rate of false alarms significantly. Although it increased the number of missed episodes slightly, from an overall view, the success of our algorithm has improved.

During the annotation studies with medical doctors in GMMA, medical doctors came across some cases for which they cannot make a decision on the class of the sound from the PSG recordings. Also two different medical doctors may label a sleeping episode differently due to their comment discrepancy. They demand extra information to be confident on their PSG results. The additional information can be supplied by our system as we can give them numerical results with probabilities. This may help medical doctors to make their decisions and maybe for their research studies.To conclude, we have reached our goals within a good level of success. Some important contributions are made with this study. We have managed to detect post apnea episodes. While doing that, we also introduced new features that are extracted from LPC error curve, and showed a different property of post apnea sounds which is the high variation of energy by calculating the entropy of energy. Also showed that they can be used for apnea detection with some improvements. Furthermore it carries additional functionalities such as being a decision aid for the medical doctors while evaluating the PSG results, which may be used in practical applications. Nevertheless, some improvements can be made to our system, which we can list as future work. First, the noise cancellation can be done by more complex methods like adaptive noise cancellation algorithms. Also to improve the success of our algorithm the database should be improved. We have worked with a limited number of data, the training data sets are not sufficient enough. We believe that with more annotated data the performance of the system will be increased. The synchronization of the sound recording system and the PSG is another important issue and should be done electronically as it will save a lot of time for the studies that will need both PSG and sleeping sound recordings as in our case. Besides making those improvements in recordings and recording set, observing the performance of our algorithm for an all-night recording should be done to in order to make our system more practical. This needs down sampling

of the recordings or recording a new sleeping database with a lower sampling rate, as it is not possible to process a 7-8 hours data recorded with 16 kHz sampling rate with the existing powerful computers in our laboratory. Although all the apnea periods do not end with a post apnea sound, the ones that end with sound may compose a new index that may be an alternative to AHI. This will be a very significant step in diagnosing apnea from only snoring sound recordings. With the previously mentioned improvements, this system can be used as a pre-screening method usable in any environment (especially patient's home sleeping environment) or a new criteria alternative to apnea / hypopnea index can be created to minimize the need of PSG.

REFERENCES

[1] Lung National Heart and Blood Institute , *Your guide to healthy sleep [electronic resource].*: U.S. Dept. of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, [Washington, D.C.] : , 2005.

[2] Shneerson J. M., *Sleep Medicine: A Guide to Sleep and Its Disorders.*: Blackwell, 2005.

[3] Teofilo M.D., Lee-Chiong, *SLEEP: A COMPREHENSIVE HANDBOOK.*: Wiley, 2006.

[4] Ivenenko A., Culebras A., *Obstructive Sleep Apnea: Diagnosis and Treatment.*: Informa Healthcare, 2007.

[5] Mikelson S. A., Fairbanks N. F., *Snoring and Obstructive Sleep Apnea.*: Lippinkot Williams & Wilkins, 2003.

[6] Ronald M., Pevernagie D., "The Acoustics of Snoring" *Sleep Medicine Reviews*, vol. 14, 2010, pp. 131-144.

[7] Berl JC, Caffier PP, "Snoring noise pollution the need for objective quanti?cation of annoyance, regulatory guidelines and mandatory therapy for snoring," *Physiol Meas*, vol. 28, pp. 25-40, 2007.

[8] Dreher A, "The annoyance of snoring.," *European archives of otorhinolaryngology*, vol. 266, pp. 293-6, 2009.

[9] Wilson J. A.,Countera P., "The management of simple snoring," *Sleep Medicine Reviews*, vol. 8, pp. 433-441, 2004.

[10] Prota R., Dalmasso F., "Snoring: analysis, measurement, clinical implications and applications," *European Respiratory Society*, vol. 9, pp. 146-159, 1996.

[11] Wen-Hung Liao, Yi-Syuan Su, "Classification of Audio Signals in All-Night Sleep Studies," in *Proceedings of the 18th International Conference on Pattern Recognition - vol 04*, 2006, pp. 302--305.

[12] Odeh M., Beck R., Gavriely N., "The acoustic properties of snores," *Eur. Respir. J.*, vol. 8, p. 2120, 1995.

13] Osman E.Z., Osborne J.E., "A new acoustic method of differentiating palatal from non-palatal snoring," *Clinical Otolarynoloy*, vol. 24, pp. 130-133, 1999.

[14] Benedetto G., Dalmasso F., "Digital processing of snoring sounds," *European Respiratory Society*, vol. 3, pp. 528-532, 1990.

[15] Mulrooney T., Wilson K., "Snoring: an acoustic monitoring technique.," *Laryngoscope*, vol. 95, pp. 1174-1177, 1985.

[16] Stone P., Agrawal S., "Sound frequency analysis and the site of snoring in natural and induced sleep," *Clinical Otolarynoloy*, vol. 27, pp. 162-166, 2002.

[17] Tassone P., Saunders NC., "Is acoustic analysis of snoring an alternative to sleep nasendoscopy?," *Clinical Otolarynoloy*, vol. 29, pp. 242-246, 2004.

[18] Mikami T., "Detecting Nonlinear Properties of Snoring Sounds for Sleep Apnea Diagnosis," in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pp. 1173 -1176, 2008.

[19] Butkov N., Lee-Chiong T., Fundamentals of Sleep Technology*, 1st Edition.*: Lippincott Williams & Wilkins, 2007.

[20] Misra H., Ikbal S., Bourlard B., Hermansky B., "Spectral entropy based feature for robust ASR," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 2004, pp. 193-196,

[21] Cushida K.,Sleep Apnea Current Diagnosis and Treatment.: KARGER, 2006.

[22] Slawinski E., Perez-Padilla JR, "Characteristics of the snoring noise in patients with and without occlusive sleep apnea.," *Am Rev Respir Dis*, vol. 147, pp. 635-644,1993.

[23] Cavusoglu M., Ciloglu T., Serinagaoglu Y., Kamasak M., Erogul O., Akcam T., "Investigation of sequential properties of snoring episodes for obstructive sleep apnoea identification," *Physiological Measurement*, vol. 29, pp. 879-898, 2008.

[24] Tuomi S.K., Duckitt W.D., "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiological Measurement*, vol. 27, pp. 1047-1056, 2006.

[25] Ng K.A., Koh T.S., Baey E., Abeyratne U.R., Puvanendran K.,.: Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? Springer Berlin Heidelberg, vol.14, pp. 967-970, 2007.

[26] Abeyratne U.R., Wakwella A.S., Hukins C., "Pitch jump probability measures for the analysis of snoring sounds in apnea," *Physiol. Meas.*, vol. 26, p. 779, 2005.

[27] Hara H., Murakami N., Miyauchi Y., Yamashita H., "Acoustic analysis of snoring sounds by a multidimensional voice program," *Laryngoscope*, vol. 116, no. 3, p. 379, 2006.

[28] Karunajeewa A.S., Abeyratne U.R.,Hukins C., "Silence–breathing–snore classification from snore-related sounds," *Physiological Measurement*, vol. 2, no. 29, p. 227, 2008.

[29] Cavusoglu M., Kamasak M., Erogul O., Ciloglu T., Serinagaoglu Y., Akcam T., "An efficient method for snore/nonsnore classification of sleep sounds," *Physiological Measurement*, vol. 28, pp. 841-853, 2007.

[30] Gold B., Rabiner L.R., *Theory and application of digital signal processing.*: Prentice-Hall, 1975.

[31] James M.D., Rowley A., Obstructive Sleep Apnea-Hypopnea Syndrome: Differential Diagnoses & Workup. [Online]. http://emedicine.medscape.com/article/302773-diagnosis, last visited on: 11/01/2011

 32] Ng A.K.,Koh T.S., Baey E., Puvanendran K., "Speech-like Analysis of Snore Signals for the Detection of Obstructive Sleep Apnea," in *Biomedical and Pharmaceutical Engineering, 2006,* pp. 99 -103.

[33] Yu Song, Wen-Hong Wang, Feng-Juan Guo, "Feature extraction and classification for audio information in news video," in *Wavelet Analysis and Pattern Recognition,* 2009, pp. 43 -46.

[34] Aye Y.Y., "Speech Recognition Using Zero-Crossing Features," in *Electronic Computer Technology, 2009* pp. 689 -692.

[35] Andrijauskas A. SoundSense. [Online]. www.eecs.ucf.edu/~lboloni/Teaching/EEL6788./T14SoundSense.pptx, last visited on 25/11/2011

[36] Saad E.M., El-Adawy M.I, Abu-El-Wafa M.E, Wahba A.A, "A multifeature speech/music discrimination system," pp. 208 - 213, 2002.

[37] Abeyratne U.R., "Pitch jump probability measures for the analysis of snoring sounds in apnea," *Physiological Measurement*, no. 26, pp. 779-798, 2005.

[38] Caruntu A., Nica A., Toderean G., Puschita E., Buza O., "An Improved Method for Automatic Classification of Speech," in *Automation, Quality and Testing, Robotics,*pp. 448 -451,2006.

[39] Theodoros G., Dimitrios K., Andreas A., Sergios T., "Violence Content Classification Using Audio Features,".

[40] Okko Rasanen, Speech Segmentation and Clustering Methods for a New Speech Recognition Architecture, M.Sc Thesis, HELSINKI UNIVERSITY OF TECHNOLOGY, Helsinki, 2007.

[41] Makhoul J., "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, pp. 561 - 580,1975.

[42] Stefanescu S., "Applying Nelder Mead's Optimization Algorithm for Multiple Global Minima," *Journal for Economic Forecasting*, vol. 4, pp. 97-103, 2007.

[43] J A., Mead R.,, "A Simplex Method for Function Minimization," *The Computer Journal*, vol. 7, pp. 308-313, 1965.

[44] Smith L., A tutorial on Principal Component Analysis, 2002.

[45] Wikipedia., Principle Component Analysis, [Online]. http://en.wikipedia.org/wiki/Principal_component_analysis, last visited on: 25/10/2010

[46] Giannakopoulos T., A method for silence removal and segmentation of speech signals, 2010.

[47] Pohjalainen J, Methods of Automatic Audio Content Classification, Licentiate's Thesis, HELSINKI UNIVERSITY OF TECHNOLOGY, Helsinki, 2007.

[48] Chen Ming Wei Jian Sheu, "AN INTELLIGENT INITIALIZATION METHOD FOR THE K-MEANS," *International Journal of Innovative Computing, Information and Control*, vol. 6, 2010.

[49] Bai L., Lao S.Y., Liao H.X.,Chen J., "Audio Classification and Segmentation for Sports Video Structure Extraction using Support Vector Machine," *Machine Learning and Cybernetics,* pp. 3303 -3307, 2006.

[50] Andersson T., Audio Classification and Content Description, M.Sc. Thesis, Luela University of Technology, Sweden, 2004.

[51] Kabre H., "Robustness of a chaotic modal neural network applied to audio-visual speech recognition," in *Neural Networks for Signal Processing,* pp. 607 -616, 1997.

[52] Xiang J.J., McKinney M.F., Fitz K., Zhang T., "Evaluation of sound classification algorithms for hearing aid applications," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 185 -188 2010.

[53] Moreno P.,Marques J., "A Study of Musical Insturment Classification Using Gaussian Mixture Model and Support Vector Machines," COMPAQ, Massachusetts, Technical Report CRL99/4, 1999.

[54] Raut C.K., Gaussian Mixture Model, Study Session on Gaussian Mixture Model, 2004.

[55] Reynolds A.D., Quatieri T, Dunn R., "Speaker verification using Adapted Gaussian mixture models," in *Digital Signal Processing*, 2000.

[56] Piater J.H., Lecture notes, Mixture Models and Expectation-Maximization, 2002.

[57] NCSSM Statistics Leadership Institute, Categorical Data Analysis, 1999.

[58] Carter N, Ulfberg J, "Adverse health effects among women living with heavy snorers," *Health Care Women Int*, vol. 21, pp. 81-90, 2000.

[59] Harris CD, Beninati W, "The effect of snoring and obstructive sleep apnea on the sleep quality of bed partners.," *Mayo Clin Proc.*, vol. 74, p. 1049, 1999

[60] Kaemingk KL, Goodwin JL, "Clinical screening of school children for polysomnography to detect sleep-disordered breathing," *Clinical Sleep Medicine*, vol. 1, pp. 247-254, 2005.