Audio Display and Environmental Sound Analysis of Diagnostic and Therapeutic Respiratory Sounds

by

Mario Garingo

Bachelor of Engineering, Ryerson University, 2011

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Audio Display and Environmental Sound Analysis of Diagnostic and Therapeutic Respiratory Sounds

Master of Applied Science 2014

Mario Garingo

Electrical and Computer Engineering

Ryerson University

Abstract

The objective of this study is to provide a framework to aid physicians in identifying early respiratory ailments as well as provide a means of monitoring medication compliancy for both the patient and physicians. To aid physicians identify abnormal sounds during auscultations such as crackle, this work proposes a multimedia approach in the form of audio display (AD) to enhance crackle sounds produced in respiration. This work utilize a two step AD approach in which the crackle sound is first separated from the rest of the vesicular sound and then either sonified or audified. To aid in monitoring use of medication this work proposes an environmental sound analysis (ESA) framework to autonomously quantify adherence to medication. This work employed traditional audio features to extract meaningful discriminatory information to identify the inhaler sounds from the environment with the aid of maximum relevance and minimum redundancy algorithm and the hidden markov model.

Acknowledgements

I would like to acknowledge my supervisors Dr. Sridhar Krishnan who has greatly helped me through out my research career, without his guidance and motivation I would not have accomplished this work.

I would also like to thank my fellow colleagues from the Signal Analysis Research (SAR) group - Mehrnaz, Farhat and Lakshmi for their guidance and company in and out of the lab.

I would also like to thank Ryan who has helped me and provided me with the equipment needed to do half of this thesis. Without his help it would not possible for me to complete this work.

I would like to thank my friends Kwame, Ramesh and Qian as well as my partner Laura for their support during my studies at Ryerson University and helping record and edit this work. Their company in stressful times was well needed to get through the though times. And last but not least I would like to thank Laura and my family who has supported and encouraged me from the very beginning to do my best and no matter the outcome of what I pursue they are proud of me.

Dedication

To my partner, Laura Muntean, my mom Lilian Garingo, and my grandmother for their continual support and encouragement, which has given me the ability to pursue my goals with confidence and full determination.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The respiratory system is one of the body's eleven biological systems and it is in charge of the exchange of oxygen and carbon dioxide between our body and the environment. As a vital system in our bodies, it is important to monitor diseases in this system as well as mitigate their impact on our health. According to the Public Health Agency of Canada (PHAC) over three million Canadians suffer from one of five respiratory related diseases[10], which include chronic obstructive pulmonary disease (COPD), lung cancer, tuberculosis (TB), cystic fibrosis, and asthma which this work will focus on. In a 2007 PHAC report, it was discovered that these chronic respiratory diseases along with others such as influenza, pneumonia, and bronchiolitis were on the rise, and that they can be linked to the increasing aging population and continuous degradation of air quality, which are two factors that we have little control over.

This increase of chronic respiratory diseases can also have a negative effect on the economy. Currently 6.5 percent of the total Canadian health care costs are used directly for respiratory related illnesses, which accounts for 5.7 billion dollars[10]. These come from hospitalization, physician visits, research and medication. Also in the 2007 PHAC report, survey results revealed that patients with asthma required immediate help in keeping their disease under control[10]. These patients had frequent emergency room visits and were often hospitalized. In addition, their quality of life declined due to activity restrictions caused by their illness.

The focus of this work will be on respiratory ailments specifically asthma, and using auditory motivated signal processing techniques to assist in early detection, intervention and mitigation of these diseases. Current respiratory sound analysis for early detection depends on highly skilled professionals who auscultate the lungs to detect abnormal sounds. One type of abnormal sound are crackles, which

have great diagnostic value as they are an indication of the early stage of a respiratory disease [11]. In the field of auscultation and crackle analysis, it is very difficult to detect crackles due to interference from various noise sources including the heart, environmental, inspiration and expiration. This work will use auditory display techniques in the context of sonification and audification to exploit the human auditory system's ability of identifying patterns, differentiating sounds, and concentrating on sound sources.

While early detection is important for respiratory disease control it is also equally important to look at patients who are currently suffering from respiratory ailments. Since the focus of this work is asthma, the current form of medication is the inhaler. If used correctly, inhalers are highly effective in mitigating asthma [12]. Unfortunately, the rate of non-adherence is 30-70 percent among asthmatic patients [13]. Beside the suffering patients endure as a result of non-adherence, this also puts a strain on the economy as it drains the health care system due to wasted medication and frequent visits to the emergency room. This work will put forward the use of environmental sound analysis (ESA) to try and create an autonomous system with the aid of mobile devices to monitor the environmental sounds and identify when a person uses inhalers. The developed system will try and quantify asthmatic patients' adherence to their medication.

The remainder of this Chapter will look at the physiological inner workings of the respiratory system in order to better understand the underlying causes of respiratory illness and their affect on the human body.

1.1 Bio-Acoustics : Applications to Respiratory System

Listening to body sounds for diagnosis, known as auscultation, has been around for the past 2000 years, but it was only in 1819 when Laennec linked the respiratory auscultation sounds to human pulmonary diseases. In 1821 he invented the stethoscope, which now enables physicians to listen for abnormal sounds and identify symptoms of diseases. With the recent advent of the digital stethoscope as well as the application of advanced digital signal processing to lung sounds, we can now denoise signals, store records of measurements, produce graphical characteristic features, extract characteristic features, as well as automatically classify lung sounds, all to help with the diagnosis and treatment of patients. Both the digital and the regular stethoscopes can be seen in Figure 1.1.

Figure 1.1: The left image is an image of a regular stethoscope [1], and the right image is an example of a digital stethoscope[2] developed by 3M.

1.2 Anatomy and Physiology: Respiratory Sounds

Before examining abnormal lung sounds, it is important to understand the respiratory system and identify sounds that are commonly heard. The main function of the respiratory system, seen in Figure 1.2, is the exchange of oxygen and carbon dioxide between the alveoli and the pulmonary circulation during inspiration and expiration. In inspiration, air enters the upper airways into the lower airway until it reaches the alveoli which are hollow cavities. Each alveolus is surrounded by multiple capillaries. During systole, the deoxygenated blood which is returning from the body's cells is pumped from the right ventricle through the arterial pulmonary circulation to these capillaries, where carbon dioxide diffuses from the blood and enters the alveolar air. Simultaneously, oxygen from the inspired air in the alveolus enters the pulmonary capillary blood. In expiration, the carbon dioxide which was diffused from the blood is exhaled from the lungs through the lower air way and out the upper airway. The oxygenated blood travels though the left side of the heart and is pumped from the ventricle into the arterial circulation to the cells of the body and the process of inspiration and expiration starts again.

The first part of respiration takes place in the upper airways and it is made up of the nasal cavities and the pharynx. The air inhaled by the body during inspiration first enters the nasal cavities, which filter out foreign particles. The lateral walls of the nasal cavities curve, forming what is called Turbinate's. This curving increases the surface areas for particles to stick to the mucus lining, as well as produce turbulent airflow. After going through the nasal cavities, the inhaled air travels to the pharynx which is made up of three sections: the nasopharynx, the oropharynx, and the hypopharynx, before entering the lower airways.

Also called the tracheobronchial tree, the lower airway starts at the larynx and ends at the distal bronchioles, which leads to the exchange surfaces of the lungs, the alveoli. The air enters the trachea which then divides into the left and the right main stem bronchus. These main bronchi enter the lung

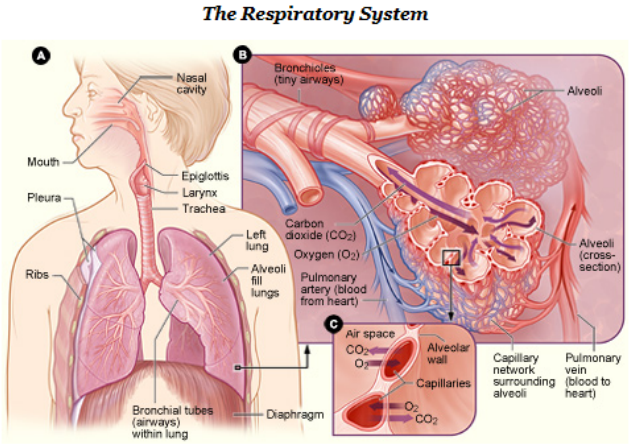

Figure 1.2: Respiratory system.[3]

at the hila, where the lung tissue attaches to the mediastinum. The bronchi branch out downward and divide into lobar bronchi which further divide into segmental bronchi. This division occurs about 25 times until terminating into the bronchioles. These bronchioles are lined with alveoli where the exchange of oxygen and carbon dioxide occurs. This process is repeated for each respiratory cycle and is considered the normal breath sound in this work.

1.3 The Mechanics of Breathing

The respiratory muscles are vital for breathing as they help the chest cavity expand and contract. This expansion and contraction is caused by the pressure difference between the atmosphere and the air in our lungs. In inspiration, the diaphragm and external intercostals muscles take action. The diaphragm flattens and the lower rib cage expands, forcing abdominal contents downward, which in turn increase the volume of the lungs. The air then flows from the atmosphere (high pressure) into the lungs (low pressure). In expiration, the thorax and the elastic recoil of the lungs increases the pressure within the lungs forcing the air out of the lower and upper airways. At the end of the cycle all the muscles are relaxed and the diaphragm reverts back to its original position.

It is important to note that the mechanics of breathing are also affected by gravity, posture and the size and shape of the lungs. When a person is oriented differently in space, gravity can cause abnormal blood distribution in the lung which causes dependent lung regions to receive a larger portion of the cardiac output than the nondependent regions. Posture effects the distribution of both air and blood flow throughout the lungs, and the size and shape effect the airflow patterns of the lung. In inspiration,

air velocity and the air rate decrease as the total cross-sectional airway areas increase. Understanding respiratory mechanics can help us understand how breathing sounds are created.

1.4 Respiratory Sounds

Breathing sounds are obtained via a stethoscope over the anterior, lateral and posterior chest wall surfaces. The breath sounds have a wide range of frequencies, but many are near the lower threshold of human hearing. As a result, current ausculation examination of the lungs occurs in quiet places so that the doctors can hear the breathing sounds associated with the inhalation and expiration during respirations, clearly and distinctly. There are two types of sounds which can be heard: normal and abnormal adventitious sounds. Normal sounds are produced by airflow patterns associated with pressure changes within the airways and by solid tissue vibrations within the lungs. Both abnormal and normal sounds are affected by the airflow patterns, regional lung volume, body position and the sound production site. These factors affect the intensity, duration, frequency and quality of the sounds being produced. Normal sounds are diminished and filtered when they are transmitted through the air-filled alveoli.

1.4.1 Normal Respiratory Sounds

Normal breathing sounds are classified as tracheal, bronchial, vesicular, or bronchovesicular. The tracheal and the bronchial sounds are produced by turbulent airflow which is loud and can be heard through the trachea and mainstem bronchi during inspiration and expiration. Vesicular sounds are faint and are best heard over the chest during inspiration and expiration. Finally, the bronchovesicular sounds are heard over the areas between the mainstem bronchi and the smaller airways. Their pitch and duration are between the tracheal, bronchial and vesicular sounds.

There are three things that affect these types of sounds: the distance between the source of the sound and the chest wall, the path of sound transmission, and the sound's location. These sounds are the loudest when heard over the trachea on the anterior chest wall surface next to the sternum. As one follows the air from the trachea to the bronchi, the normally high-pitched sounds become diminished due to the filtering behaviour of the chest wall, pleurae, and air-filled lung tissue. However, over most chest walls, the normal breathing sound is soft and low pitched. Note that normal sounds are softer and shorter during expiration than during inspiration.

1.4.1.1 Tracheal and Bronchial Respiratory Sounds

Tracheal and bronchial sounds are produced by turbulent airflow patterns which are heard over the trachea (tracheal sounds) and mainstem bronchi (bronchial sounds). Tracheal sounds are harsh and high pitched, while bronchial sounds are loud and high pitched. There is a pause in the end of inspiration and expiration is longer than inspiration (I:E ratio of about 1:2 to 1:3). The frequency of the sounds is found in the 200-2000Hz band.

1.4.1.2 Vesicular Respiratory Sounds

Vesicular sounds are transmitted through lung tissues and the chest walls, and are produced by changes in airflow patterns. These sounds tend to be quieter than the tracheal and bronchial sounds. Inspiration can be clearly heard, which is followed by a short expiration due to the decline of the airflow and the upward directed flow of the turbulent airflow to the central airways. The I:E ratio ranges from 3:1 to 4:1. The frequency of the sounds is found in the 200-600Hz band.

1.4.1.3 Bronchovesicular Respiratory Sounds

The intensity of the bronchovesicular varies depending on where the sound was acquired as well as the posture of the patient. These variations are directly related to the airflow patterns and the distribution of air within the lungs. Many of these changes may not be audible . Typically, the brochovesicular breath sounds have a low pitch and are softer and less harsh than the bronchial breath sounds. The bronchovesicular sounds have a tubular quality and have an I:E ratio of 1:1.

1.4.2 Abnormal Respiratory Sounds

As the name suggests, these sounds are usually out of the ordinary and stand out from what is normally heard. These sounds are categorized as either being caused by turbulent airflow such as in the case of bronchial breath sounds or an additive nature in the case of adventitious sounds.

1.4.2.1 Bronchial Respiratory Sounds

Abnormal bronchial breath sounds are mainly caused by an increase in lung density which can be caused by consolidation , atelectasis , and fibrosis . The increase in density is caused by fluid accumulation, lung collapse, or fibrotic scarring. The increase in density means that the transmission of breath sounds from large airways is enhanced. In other words, due to the increased impedance between the fluid-filled lung tissue, the pleurae and the chest wall, there is a decrease in high frequency sounds. The increase

in lung density therefore causes the transmission of these sounds more readily to the chest wall surface which makes the sound louder and last longer than normal. The I:E ratio changes to 1:1 or 1:2 from 3:1 or 4:1.

1.4.2.2 Adventitious Respiratory Sounds

Adventitious sounds are added sounds that are heard with the normal breath sounds. These sounds are either crackles or wheezes which are then further classified by duration and pitch. Crackles and wheezes are sometimes called discontinuous and continuous respectively. This is because crackles are very short bursts of sounds which are perceived as discontinuous, and wheezes have a musical aspect to them and are perceived as continuous sounds.

Crackles sounds are characterized by short, explosive, popping sounds, which are classified according to their pitch, timing and location. Depending on their underlying cause, the characteristics of these crackles change. There are two main generators of crackles: air bubbling through secretions and the sudden, explosive opening of airways. Discontinuous explosive sounds are loud and low pitched and are sometimes called coarse crackles, rales, or coarse rales. Discontinuous explosive sounds that are shorter in duration, higher in pitch and less intense than coarse crackles are called fine crackles, fine rales or crepitations.

Wheezes are caused by the flutter and oscillation of the lung walls over narrow airways. They are heard in both inspiration and expiration, and will influence the wheezing sounds. Wheezing will make expiration seem longer than usual and inspiration will sound less pure. Continuous sounds that are high pitched and have a hissing or coughing characteristic are called wheezes. Continuous sounds that are low pitched and have a snoring characteristic are called low-pitched wheezes and are sometimes called sonorous rales, sonorous rhochi. These sounds are caused by fluid or secretions which are blocking large airways and they change in sound or disappear when the patient is coughing.

Similar to wheezes, Stridors are musical continuous adventurous sounds. They are mainly caused by a partial obstruction in a central airway near the larynx. Squawks are also very similar to wheezes but only occur in allergic situations and interstitial fibrosis. These sounds are usually following a crackle and are caused by explosive opening and fluttering of the unstable airways, which then produces short wheezes.

1.5 Asthma

The Asthma Society of Canada (ASC) states that asthma affects over three million Canadians while affecting over 235 million people worldwide [14]. The ASC have defined it as chronic inflammatory disease of the airway. Being a chronic disease means it has to be monitored and controlled over the duration of a persons life time. The cause and cure of asthma is currently unknown, but medication in the form of inhalers can reduce the symptoms. Symptoms include shortness of breath, tightness in the chest, coughing and wheezing. These symptoms are triggered by many factors but are usually caused by particles entering the airways. Regardless of whether a person has asthma or not, when particles enter the airway, the air flow is obstructed as it travels in and out of the lungs, which is explained in detail in Section 1.2. For most people, these particles will simply invoke coughing which dislodges these particles, but in the case of asthmatic patients the lining of the airways become more inflamed and may produce more mucous, which causes sensitivity and induces an asthma attack. Due to the increased sensitivity, muscles that surround the airways can also start to twitch and tighten; reducing the airways which further increases the severity of an asthma attack.

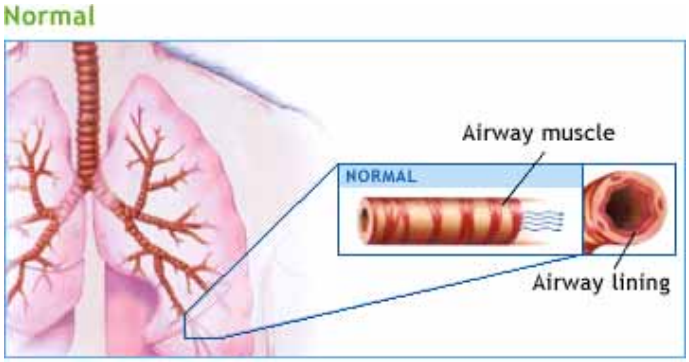

Figure 1.3: Normal airways during respiration.[4]

There are currently two main issues concerning mitigating asthma: diagnosis and adherence to medication. In terms of diagnosis, asthma has several symptoms but can be misdiagnosed with other respiratory ailments such as viral or sinus infections. This is because symptoms can vary from person to person, ranging from mild to moderate to severe, and they can randomly occur from time to time and then not appear again for long periods of time. Due to the complexity of this issue in the mitigation of asthma, this work will not focus on this issue but rather on the issue of adherence to medication.

Inhalers are currently the most effective method to combat and reduce asthma attacks from occurring because the medication is inhaled and is applied right to the source. Other medication types such as pills will first need to be ingested and then through the bodys metabolism travel to the lungs, via the blood

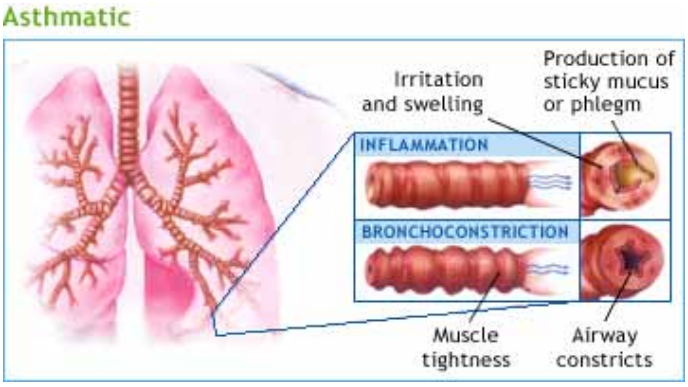

Figure 1.4: Abnormal airways in asthmatic patients.[5]

stream, where it is needed. Therefore, an inhaler is more desired compared to other type of medications. Inhalers fall under two main types: aerosol and dry-powder inhalers. Aerosol inhalers are pressurized canisters which contain medication. When the canister is pressed, a dose of medication is propelled through the inhaler opening and is inhaled by the patient. Dry-powder inhalers contain dry powder which is inhaled and drawn to the lungs. If used properly, both types of medication can control asthma attacks and therefore choosing medication is up to the patient as it is based purely on preference. Some use dry powder because it is easier to coordinate breathing and the inhalation of the medication, while others prefer aerosol because they do not need to breathe in that heavily. For the purpose of this work we will focus on aerosol inhalers as they create a high pressurized sound which this work will try and detect in various sound environments.

Figure 1.5: Aerosol inhaler example.[6]

The main issue with the current form of medication is the adherence to the medication and the

misuse of the device by the patients. Both of these issues are caused by the lack of knowledge transfer from doctor to patient. There are two types of medication: the preventer type and rescue type. The rescue type of medication is the one everyone is most familiar with. This type is used when a person is having an asthma attack - they simply use their inhaler to stop the attack from continuing. On the other hand, the preventer type medication, the least commonly known type and the focus of this work, is used several times a day even without an asthma attack occurring. This is to ensure that the airways stay clear and prevent swelling or mucous build up. The problem occurs with people who are using the preventer type medication. Patients are under the assumption that they only need to use inhalers when an attack arises, but in reality patients need to be using their inhalers regardless. This lack of adherence to medication causes many asthmatic patients to be hospitalized and work such as [13][15][16] show ways to reduce these emergency visits to the hospital in terms of properly educating the patients about the use of inhalers. Furthermore, this lack of adherence can result in poor clinical response to the medication causing patients to believe that it does not work and further worsens their condition.

Adherence to medication can be broken down to temporal and technique adherence [17]. Technique adherence refers to using inhalers properly, while temporal adherence refers to the amount of time the patient needs to be using the inhalers per day. This work will focus on temporal adherence. In spite of the cause of adherence, the results are still the same: wasted medication resulting in higher health costs and higher mortality rates.

This work will present a method to automatically monitor a patients temporal adherence without the patients knowledge through an Environmental Sound Analytics (ESA) framework [18]. The main objective is to track their adherence to the inhaler medication and give patients feedback, ultimately encouraging patients to take action on their poor adherence and improve their quality of life while reducing their emergency visits. To achieve this a smart phone app will be used to record environmental sounds continuously and identify when a patient has used their inhalers.

1.6 Anatomical and Physiological: The Human Ear

The engineering approach used in this work will use techniques either derived from or that employ the human ear. Hearing is one of our five senses and it has been honed through evolutionary means to discern sounds in the environment. It is extremely sensitive and if properly trained, can separate individual sounds in a conglomerate recording of sounds. This high recognition capability of the human ear is a very a useful trait which both ESA and AD can utilize to obtain better results.

The main purpose of our ears is to convert sound pressure or waves into nerve impulses which can

then travel to our brains for processing. This process starts in the vibration of the tympanic membrane which are made up of the many small bones in the inner ear. These vibrations create small pressure variations within the fluid filled canals (called the cochlea) which are sensed by nerve endings (both the vestibular and the cochlea) which transfer them to the acoustic nerve and then to the brain. The auditory system acts like a computer storing past experiences and allows humans to focus on sounds of interest within different noise conditions. This is the reason people are able to have a conversation even in loud noisy city areas. This work explores features which try to emulate and exploit our understanding of the inner workings of the human ear to get a better picture of audio signals.

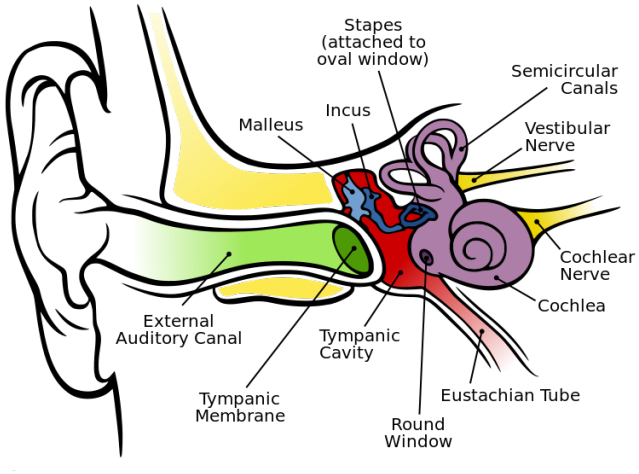

Figure 1.6: Human auditory system.[7]

The frequency range in which humans can hear is called the audible frequency range and it spans from 16 Hz to 20 kHz. The threshold of hearing ranges from 0 db to 130 db; anything more than 130 db and we start to experience pain. Our ears are mostly sensitive to the sounds that are in the range of 500 Hz to 4 kHz, which encapsulates the frequencies of normal speech. Lower frequency sounds are very difficult to hear and need larger amplitudes in order for us to hear them. This is important to note because normal lung sounds are mainly below 400 Hz.

Certain characteristics of sounds such as pitch can also alter our perception of sounds. The pitch of a sound can be easily heard in pure tones which last for one tenth of a second or more. Anything under will result in a decrease in perception of intensity and will appear to sound like clicks. Therefore the duration of the sound will greatly influence our perception of its pitch.

1.7 Database and Data Acquisition

1.7.1 Digitizing Audio Signals

Before delving into manipulating and extracting features from signals, it is important to get an understanding of the data acquisition procedure used in this work. In general, real world sounds have two properties, they have frequencies and volume. In contrast, the digital domain sounds have sample rates and bit rates. The sampling rate determines the descriptions of the analog frequencies in the digital domain while the bit rate describes the volume in the digital domain. These two properties, frequency and volume describe a sound both in the analog and digital domain and are needed to fully describe the signal.

Sampling rate can be seen as how often the computer describes a continuous analog signal. Keeping this in mind, based on the Nyquist-Shannon sampling theorem, the sampling rate has to be twice as much as the highest frequency of the analog signal. This means if the highest frequency of the analog signal is 10kHz the sampling rate must be at least 20kHz in order to describe the full spectrum of the analog signal in the digital domain. Since humans can only hear from 20Hz to 20kHz, we typically choose a sampling rate of 44.1kHz which can capture frequencies up to 22.005kHz. There are three other possible bit rates: the ultra-wideband which samples at 32 kHz, the wideband which samples at 16 kHz and finally the narrowband which samples as 8kHz.

Similar to the sampling rate, bit rates sample the continuous signal, but instead of sampling the signal in the time domain, they sample the volume of the signal, describing the volume in a set range. Different ranges will determine the amount of volume that is described. The higher the bit rate, the more accurate we can determine exactly how loud or quite the volume is in the analog domain.

1.7.2 Auscultation Sound Data

The R.A.L.E database [19] was used in this work for data pertaining to respiratory sound analysis. The database has several recordings for various normal and adventitious sounds including wheezes and crackles. The recordings were in a two channel stereo 16-bit wav format ranging in size from 440kb to 500 kb.

In the clinical setting auscultation does not simply mean listening into the respiratory sounds. Due to the complex sound propagation through the lung, chest and skin, sounds vary depending upon the location in which auscultation is taken. It is recommended that auscultations need to be taken on eight locations on the anterior chest and ten locations on the posterior chest [8]. These locations can be seen in Figure 1.7. It is also recommended that auscultations must be done on opposite sides of the thorax

successively in order to compare left and right lung sounds as well as it creates a systematic procedure in which to accurately identity the location of the abnormality for a more accurate assessment of ailments. Furthermore, ausculations must be done on the patient's bare skin (to reduce the sound filtering nature of clothes), and asked to breathe through the mouth rather than through the nose (to maximize air flow going into the lungs)[8].

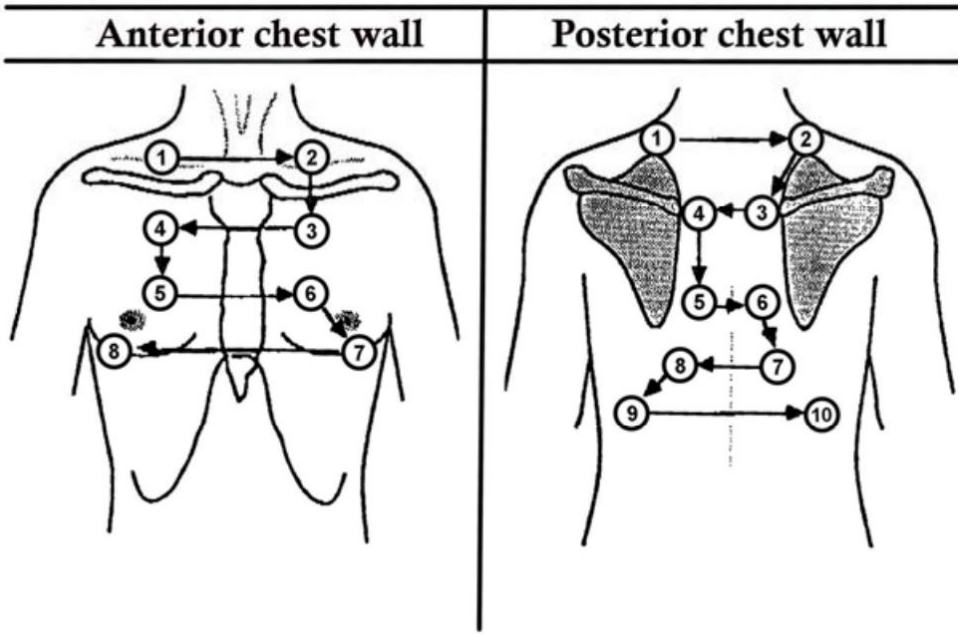

Figure 1.7: Human auditory system.[8]

1.7.3 Inhaler Sound Data

To our knowledge, this is the first time anyone has ever tried to do this type of asthma medication compliance quantification through the ESA framework. As such, there are no publically available databases to perform tests on. Consequently, in this work, a small database was generated to test the proposed ESA scheme. Over 1000 sound samples were captured in various settings and various distances from the microphone were captured to create the database.

In order to get a comprehensive test on how well the system fairs in the real world, inhaler sounds were recorded (through the method described on eariler in this section in various environments which range from quiet places, such as the library, to loud places, such as the subway. In addition, because the real world application of this system is to be implemented on a mobile device, it is also important to record the inhaler sounds in multiple distances from the microphone. This is because mobile devices

are seldom on a fixed distance from a person's mouth (i.e. when their using an inhaler).

Recording was done using audacity which is free software that can be used to manipulate many aspects of the recording such as bit rate and sampling rate. For our purpose, we will choose a sampling rate of 16kHz instead of 44.1kHz to try and reduce the noise components of the surroundings.

1.7.3.1 Distance Measurements

The distance from the microphone to the inhaler is a variable that will be discussed in this section. Distance is a factor in identifying the features of the inhaler sound since the sound characteristics may change as the distance from the inhaler to the microphone changes. In our scenario, we are considering that the microphone will mostly likely be in a person's pocket. Therefore using publically available human average measurements from NASA (see Figure 1.8) the maximum distance from the mouth to the pocket is about 2.5-3 feet or approximately 75-90 cm. Therefore, in the worst case scenario, the farthest distance from the recording device to the inhaler will be about 90 cm. In this work, we gradually increased the distance from the inhaler to the microphone from 10 cm to 90 cm in 10 cm increments and recorded the raw inhaler sound in various environments.

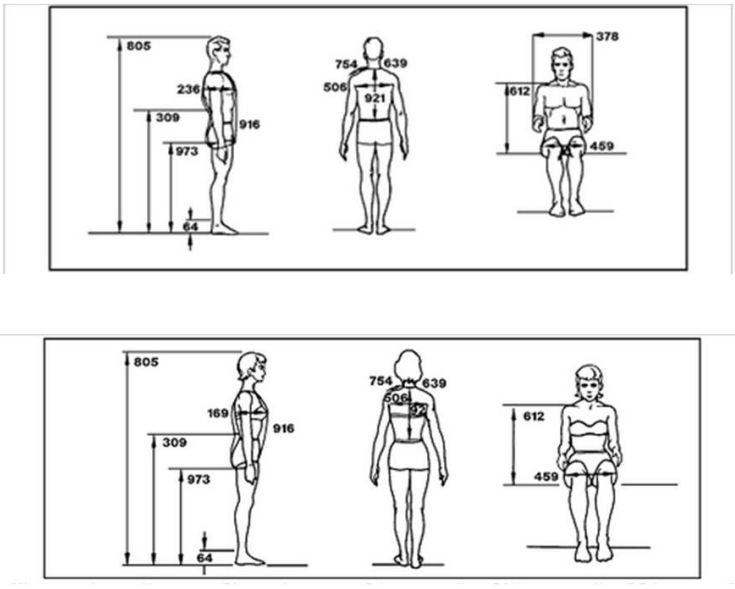

Figure 1.8: NASA average male and female measurements.[9]

1.8 Thesis Contribution

There are two main contributions of this work which are expanded upon on chapters two and three. The first novel contribution is the application of AD to a respiratory sounds. Though AD have been applied to biomedical signals before, such as the brain, the heart and the knee. Based our literature survey, this is the first time it has been applied to respiratory sounds. This novel AD approach to respiratory sounds, introduces two unique stages seen on Figure 1.10. The first stage consists of the separation stage which separates the wanted signal from the unwanted signal. This separation is needed to emphasize only the meaningful information. Due to the subjectivity of sounds the second stage enables a user to select either to apply audification or sonification. The audification is prior art but the sonification mapping is unique to this work. This work utilizes three known TF representation, and maps them to a MIDI paradigm in which pitch, velocity, duration and timbre is controlled by the TF map as well as the energy of the signal.

The second contribution is the application of ESA to autonomously quantify medication adherence of asthmatic patients. Though ESA has been on the rise due to its security and smart home application potentials, it is a fairly new branch of audio signal processing and has only been applied to very few biomedical application. As such this work tries to utilize the ESA to identify when an asthmatic patient uses their inhaler to quantify and monitor adherence. Though the general approach is prior art in-terms of feature extraction, feature reduction and classification. The features selected in the proposed framework try to utilize some acoustic features to try and describe non-acoustic sounds. Another contribution that stems from this work is the creation of a database consisting of over 2000 sounds segments, which have been recorded and labeled in various environments.

1.9 Thesis Organization

As previously stated, this work investigates the possibility of using AD and ESA along with auditory inspired techniques, to help in early detection and mitigation of respiratory ailments, specifically asthma. In the field of early detection of lung disease using auscultation of the lungs, AD can be used to enhance abnormal sounds, which are currently hard to detect using current auscultation techniques. The goal is to perform sonification and audification techniques on the separated abnormal sounds from regular respiration sounds, and manipulate them in such a way that we can employ the discriminatory power of the human auditory system. In the field of mitigation of lung diseases, this work will look at the adherence to asthma medication through the use of ESA techniques to automatically identify when a person uses their inhalers. This is done by using environment recordings through the use of mobile

devices, and identifying inhaler sounds based on their signature characteristics. The remainder of this thesis is organized as follows:

Chapter 2, Signal Analysis, is a literature review of techniques for separating abnormal sounds from normal sounds, different types of auditory displays (AD) and environmental signal analysis (ESA). AD and ESA .will be applied in the context of respiratory ailments to assist in early detection and mitigation of these diseases as these techniques have not yet been applied to this context.

Chapter 3, Audification and Sonification of Breath Sounds, will detail the techniques used in the proposed AD system to enhance crackle adventitious sounds in respiratory recordings as well as their ability to intensify crackle sounds to discern different crackle types in a lung recording.

Chapter 4, Inhaler Detection Based on ESA Techniques, introduces an ESA system to automatically detect inhaler sounds from the environment, creating a system to monitor patient medication adherence to inhalers.

Finally, Chapter 5, Conclusion and Future Work, provides an overview of the proposed algorithm and future improvements.

The flowchart Figure 1.9 below illustrates the flow of the remaining chapters of this thesis with respect to its contribution. The main contribution of this thesis is the application of AD and ESA to respiratory based signals. The flowchart of the proposed methods for each can be seen in Figure 1.10 and Figure 1.11 respectively and the highlighted regions are where the contributions of this work lie.

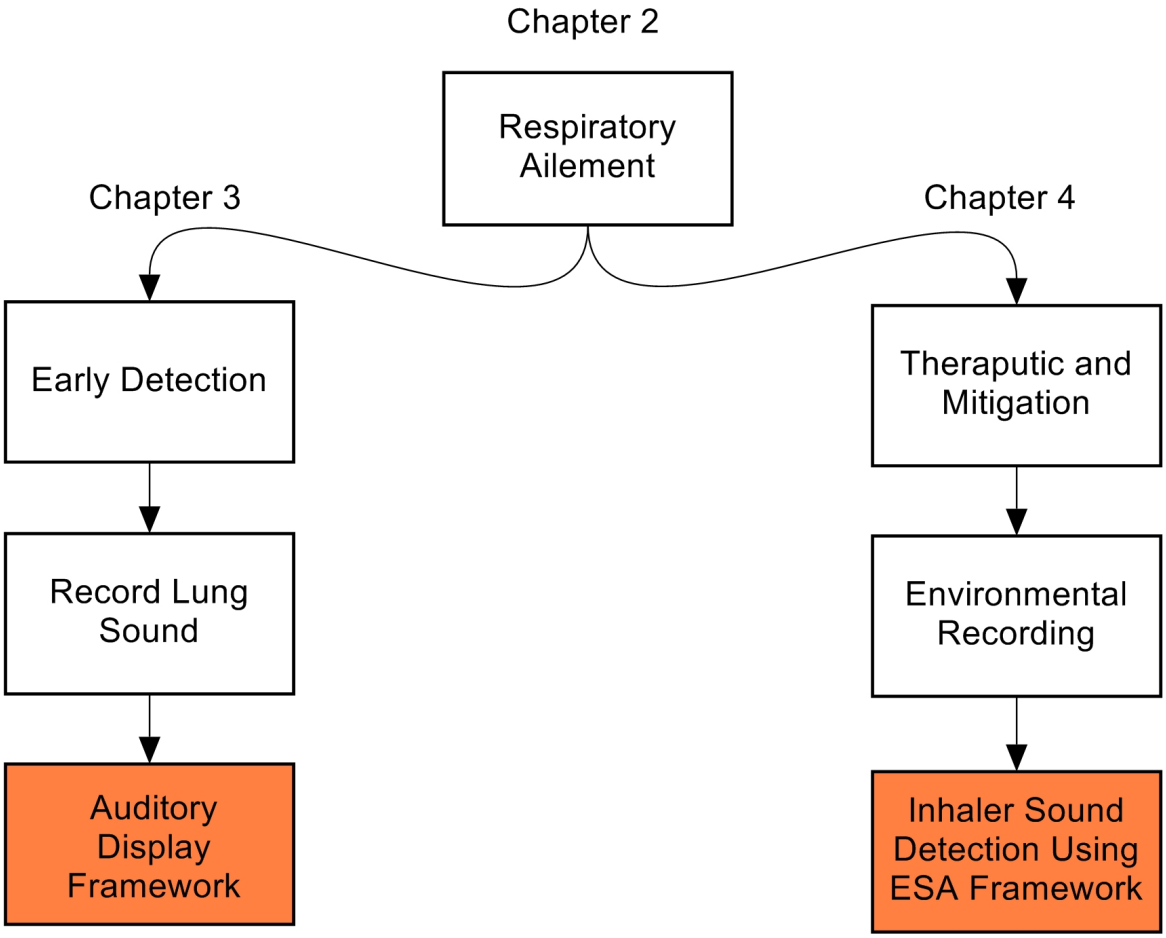

Figure 1.9: Flowchart of chapters which highlight thesis contribution.

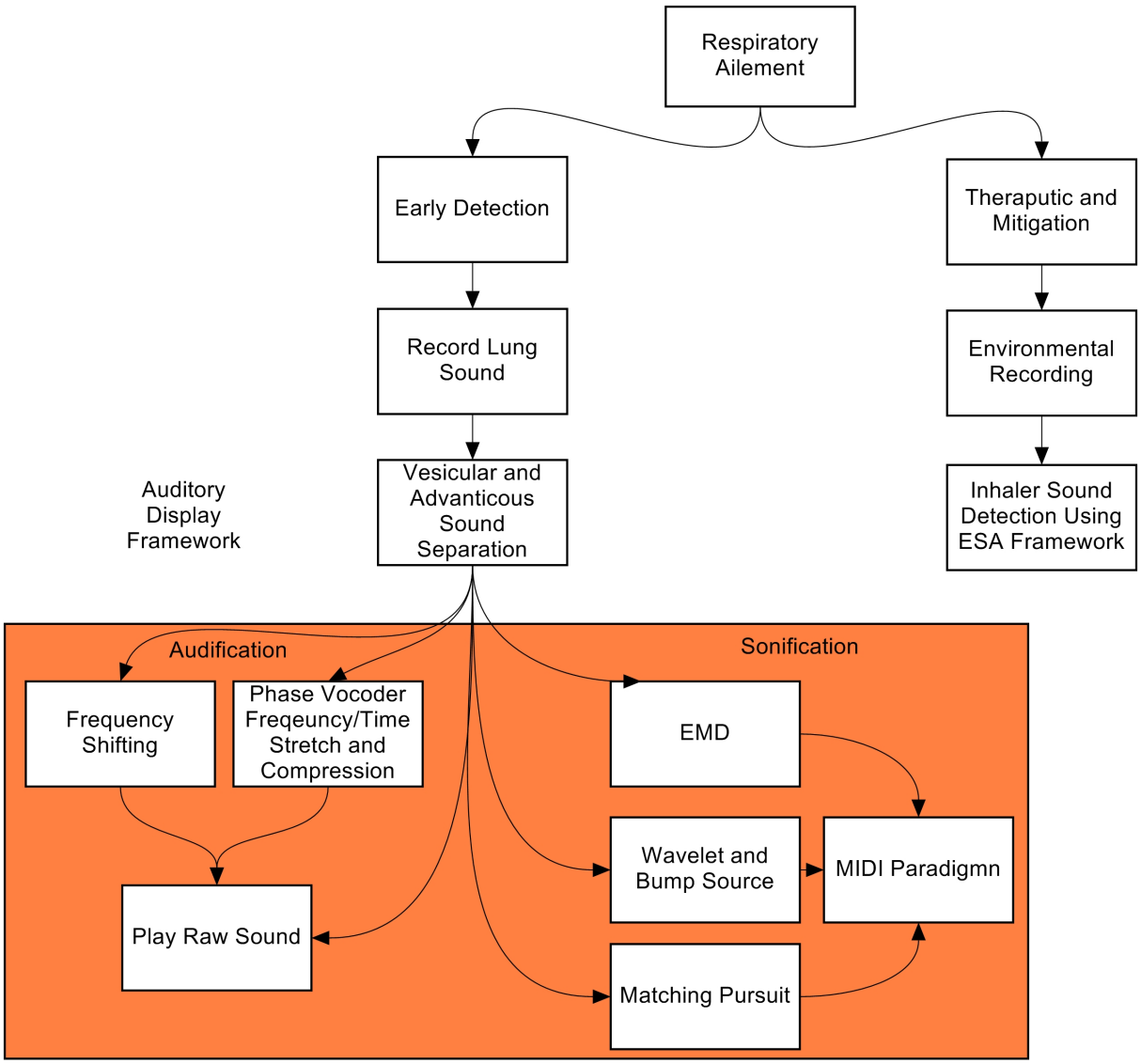

Figure 1.10: Expanded flowchart of the proposed auditory display framework for early detection of respiratory ailments, using both sonification and audification.

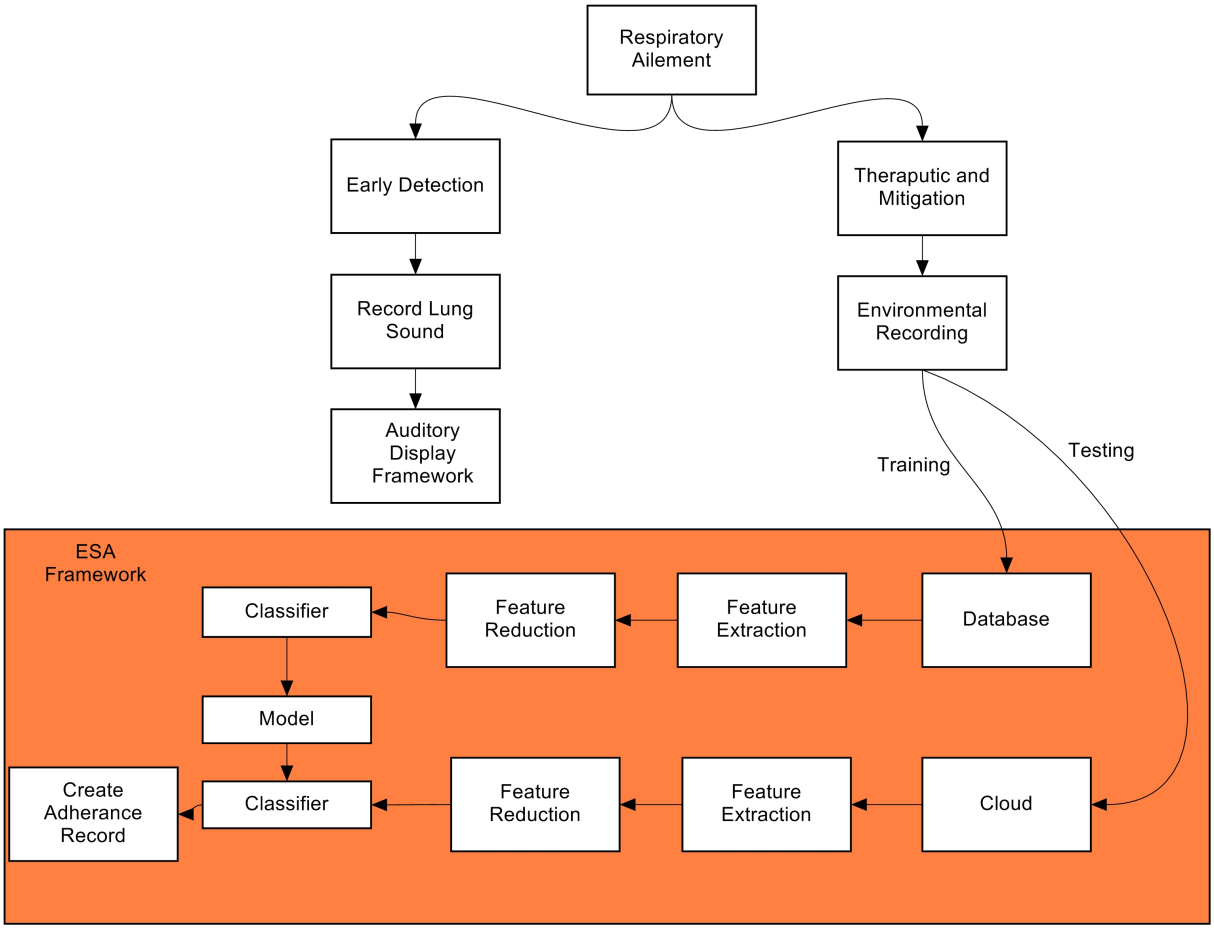

Figure 1.11: Expanded flowchart of the proposed environment signal analysis for inhaler usage detection for the assistance of mitigating asthma.

Chapter 2

Signal Analysis

There are many auditory displays (AD) and environmental signal analysis (ESA) algorithms that have been currently established and used for both non-biomedical signals and biomedical signals. This chapter explores some AD and ESA techniques which can be used for respiratory sound analysis. In the case of AD, the main purpose is to enhance the underplaying signal characteristics that are useful for early detection, such as crackles. In order to do that, sonograms which are currently being used in practice, must be replaced with more advanced signal processing techniques such that one can identify the explosive characteristics of crackles more efficiently. This work will look at Wavelet, Empirical Mode Decomposition (EMD), and sparse techniques to tackle both the time and frequency domain to detect crackles, and audification and sonification to enhance them. Wavelet, EMD and sparse techniques have been used for respiratory sound analysis before but to our knowledge this work may the first to apply AD techniques to enhance the separability of the crackles from the other respiratory sounds. There are many AD techniques that have been developed in the past but this work explores audification and sonification which are explained in this chapter.

ESA is a new branch of audio processing that tries to recognize particular natural and artificial sounds that are useful and need characterising in the environment. To our knowledge, this work may be the first to apply this technique to try and quantify asthma medication adherence autonomously by use of a microphone to detect inhaler made sounds. There are many ESA applications and each application has a catered feature set, and some features may be more meaningful than others. We will explore current ESA feature extraction techniques specifically those that model the human auditory system such as the Mel and Gammatone frequency cepstral coefficients. This work will also utilize a feature reduction technique called maximum relevance and minimum redundancy (mRMR) to perform

feature selection, and finally we look at the commonly used classification method hidden markov model to perform discrimination from the inhaler made sounds from the environmental sounds. These ESA techniques will be looked at in this chapter.

This chapter is divided into three sections. The first section investigates signal processing techniques that are commonly used for crackle detection in breath sounds, and auditory display techniques to be applied to the modified breathe sounds for enhancements. The second section explores different ESA feature extraction techniques, feature selection techniques and classification techniques for autonomous inhaler sound detection for asthma adherence. Finally, the third section briefly explains the idea that both the crackle and inhaler sounds are essentially noise (in terms of its audible characteristics) and that both the ESA and AD techniques explored in this work are essentially noise enhancement and noise detection techniques.

2.1 Signal Processing Techniques For Respiratory Sounds

The acoustic characteristics of various breathing sounds are caused by different sources or transmission paths of the sounds inside the lungs. This suggests that these anatomical changes will ultimately be reflected in the characteristics of their time and frequency domain counterparts. The main focus of this section is to identify signal processing techniques which can highlight the characteristics of the breathing sounds which can then be further used to discriminate abnormal sounds from the normal sounds.

Many researchers have applied both linear and nonlinear digital signal processing techniques to try and isolate adventitious sounds from the recorded lung sounds. Adventitious sounds, as described in the previous chapter, are categorized into two categories: crackles and wheezing. Crackles are explosive bursts of discontinuous sound while wheezes are more musical continuous sounds which are super imposed on the normal lung sound. Between these two categories, the crackle adventitious sounds have been widely examined as they are more important because of their diagnostic value in early detection of illnesses such as asthma. As such, this work will focus on crackle sounds.

This section will highlight three advanced signal processing techniques to try and obtain separation of crackle sounds from the lung sounds. These three signal processing techniques are wavelet, sparse representation and empirical mode decomposition (EMD). By separating abnormal sounds (crackles) and normal lung sounds, we can then apply audification or sonification techniques to enhance the physicians ability to diagnose the patient and try to reduce the subjective nature of auscultation.

2.1.1 Spectrogram

Currently spectrograms or sonograms of the lung sounds are being used by physicians to aid them in analyzing the lung sounds in computer aided auscultations. Spectrograms are power spectrums for each time segment of the lung sound. The horizontal axis is time while the vertical axis is the frequency component of the signal. The colour represents various strengths in the power spectrum. The following images are the spectrogram of the most common lung sound and adventitious sounds.

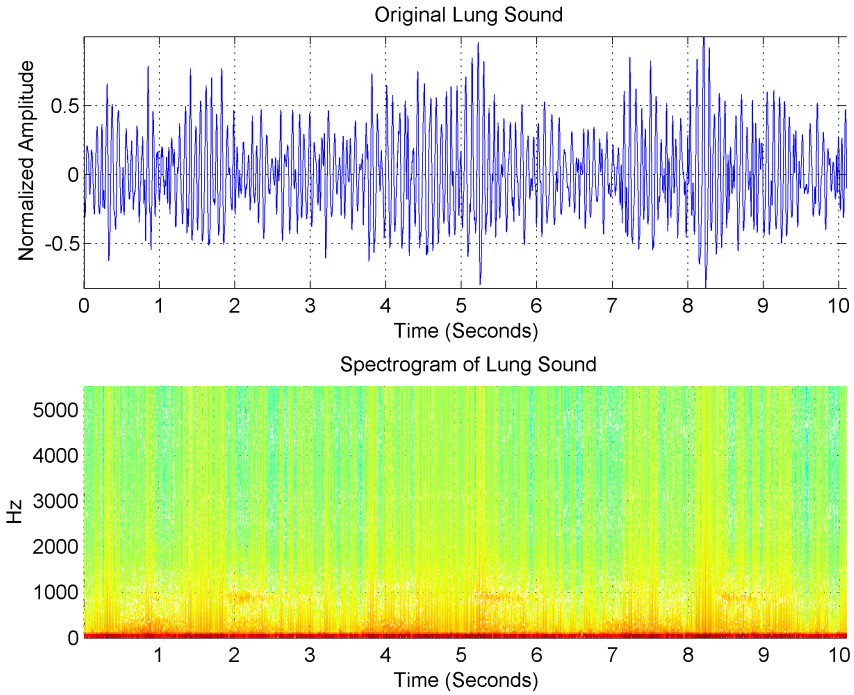

Figure 2.1: Spectrogram of a normal bronchial sound.

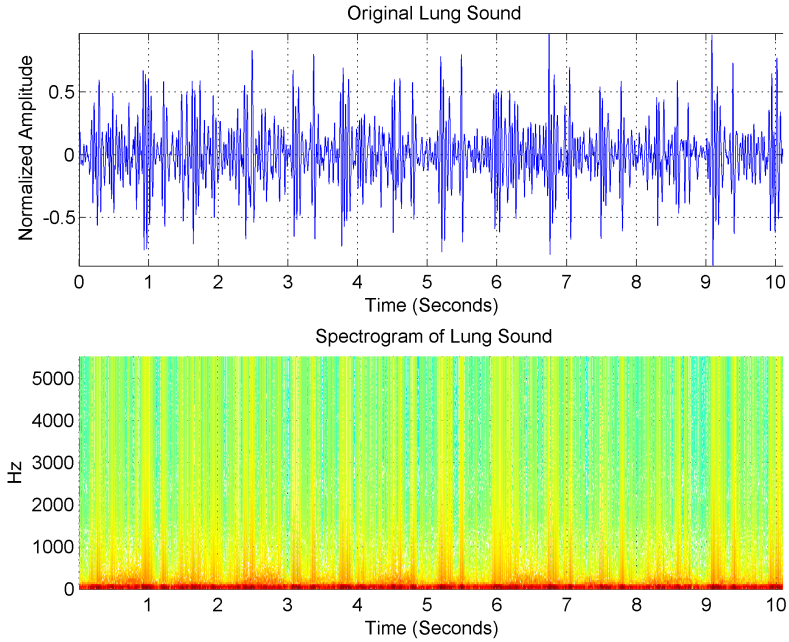

Figure 2.2: Spectrogram of a normal vesicular sound.

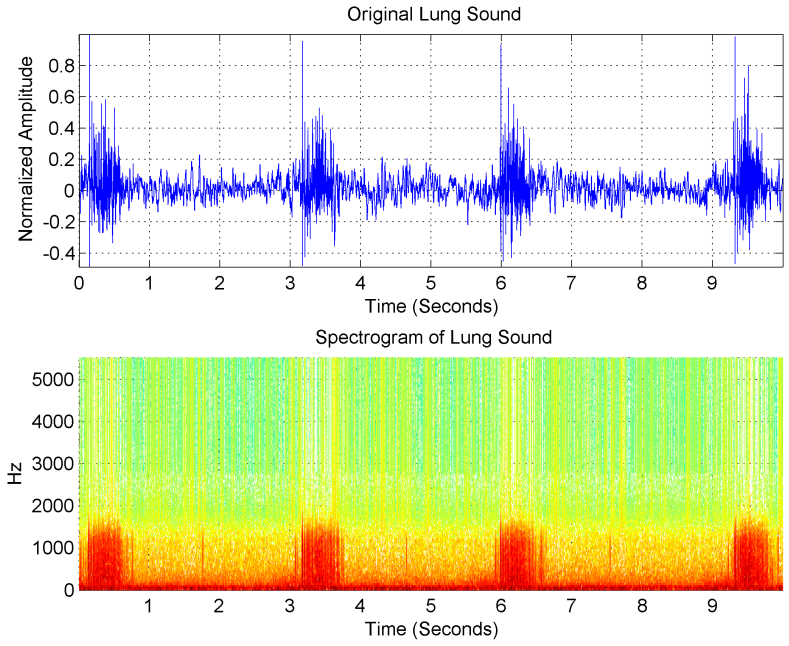

Figure 2.3: Spectrogram of a crackle adventitious sound

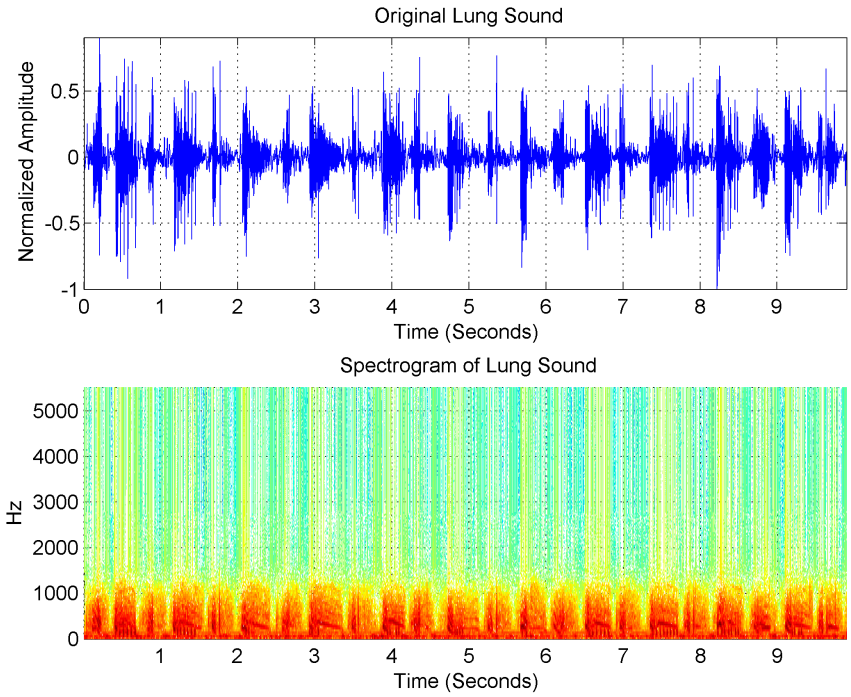

Figure 2.4: Spectrogram of a wheezing adventitious sound of the lung.

2.1.2 Wavelet

The premise behind using wavelet as a separation tool of crackles from lung sounds stems from the explosive nature of the crackle. This explosive characteristic means that the crackle has a transient characteristics in the time domain and broadband energy on the frequency representation. Therefore a time-frequency (TF) based analysis is needed to achieve proper analysis of the signals. Based on the multi-resolution decomposition as well as its simultaneous analysis of the time and frequency domain, wavelet packet transformation (WPT) can be useful in achieving separation of lung and crackle sounds.

The wavelet packet transformation is an extension of the wavelet transformation which offers more flexibility in dealing with various kinds of signals. Similar to the wavelet transform, there are two types of Finite Impulse Response (FIR) filters, high and low, which decompose the signal into multiple scales in WPT. But unlike wavelet transform, WPT iterates the signal through both the high and low pass filters. This means that the more decomposition levels the algorithm goes through, the more complex the search for a basis function becomes. Therefore an entropy cost-based criteria is used to find the best basis function of the WPT and use the basis function to reconstruct the signal.

During this separation process the crackles are extracted by two thresholds of the WPT coefficient

at each decomposition level to differentiate the lung and crackle sounds. The separated coefficients are reconstructed via best basis selection and the inverse wavelet packet transforms (IWPT). The algorithm is as follows;

1. Split signal into a short length (L=1024) and overlapped (75 %) segments
2. Apply the WPT to each segment.
3. Score the coefficients $W_k^j[m]$ whose amplitude exceeds the threshold $S1^j$ which is defined by: $S1_k^j = P_1\sigma_k^j$ where σ_k^j is the standard deviation of the wavelet coefficients corresponding to the kth subband of the level j. The coefficient P_1 is empirically fixed to 0.75. Thus, the scoring $M_k^j[m]$ of one coefficeint $W_k^j[m]$ is defined as $M_k^j = \binom{1 if W_k^j[m] \geq S1_k^j}{0 else}$
4. Quantify the number of scoring related to each coefficient $W_k^j[m]$ for all subbands k of the level j: $N^k[m] = \sum_{k=1}^{2^j} M_k^j[m]$. Define a second threshold $S2_k^j$ related to the mean number of scoring of the level j: $S2^j = P_2 \frac{1}{\frac{L}{2^j}} \sum_{m=1}^{\frac{L}{2^j}} N^j\lceil m \rceil$. The coefficient P_2 is fixed experimentally to 2.
5. Separate the coefficients $W_k^j[m]$ in two classes: stationary $S\_ST_k^j[m]$ and non-stationary $S\_NST_k^j[m]$ according to the values of the thresholds $S1_k^j$ and $S2_k^j$.

   Initialization : $S\_ST_k^j[m] = 0$; $S\_NST_k^j[m] = 0$;
   if $W_k^j[m] \geq S1_k^j$ and $N^j[m] \geq S2_k^j$, then

   $S\_ST_k^j[m] = W_k^j[m]$
   else
   $S\_NST_k^j[m] = W_k^j[m]$
6. Select the best basis tree from the WPT coefficients $W_k^j[m]$,
7. Each segment is divided by the number of overlapping times. Stationary and non-stationary signals are obtained with the inverse wavelet packer transform (IWPT) of the corresponding coefficients $S\_ST_k^j[m]$ and $S\_NST_k^j[m]$

2.1.3 Empirical Mode Decomposition

It has already been established that when examining crackle sounds within the mixture of lung sounds, TF analysis is needed to fully differentiate between the sounds. But the non-stationary behaviour

of crackles has not been discussed. Empirical Mode Decomposition (EMD) is an algorithm proposed by Huang et al. which decomposes the signal into components which have well-defined instantaneous frequencies. Each oscillatory mode that is extracted by the algorithm, known as Intrinsic Mode Functions (IMF), are symmetric, they have a unique local frequency, and each IMF exhibits a different frequency at different times. EMD is data driven in that the basis of the expansion is of an adaptive nature confined to the data itself. This implies that all data can be decomposed based on its simple intrinsic mode oscillations (mainly IMFs). The algorithm can be described below.

1. Identify the successive extrema of $x(t)$ based on the sign alteration across the derivative of $x(t)$.
2. Extract the upper and lower envelopes by interpolation, that is, the local maxima (minima) are connected by a cubic spline interpolation to produce the upper (lower) envelope. These envelopes should cover all the data between them.
3. Compute the average of upper and lower envelopes $m1(t)$.
4. Calculate the first component $h1(t) = x(t) - m1(t)$.
5. Ideally, $h1(t)$ should be an IMF. In reality, however, overshoots and undershoots are common, which also generate new extrema and sift or exaggerate the existing ones. To correct this, the sifting process has to be repeated as many times as is required to reduce the extracted signal as an IMF. To this end, treat $h1(t)$ as a new set of data, and repeat first to fourth steps up to k times (e.g., k = 7) until $h1k(t)$ becomes a true IMF. Then set $c1(t) = h1k(t)$. Overall, $c1(t)$ should contain the finest scale or the shortest period component of the signal.
6. Obtain the residue $r1(t) = x(t) - c1(t)$.
7. Treat $r1(t)$ as a new set of data and repeat first to sixth steps up to N times until the residue $rN(t)$ becomes a constant, a monotonic function, or a function with only one cycle from which no more IMFs can be extracted. Note that even for data with zero mean, $rN(t)$ still can differ from zero.
8. Reconstruction can be done using the following equation: $x(t) = Ni = 1ci(t) + rN(t)$.

Due to mode mixing, where frequencies are present in multiple IMFs, it is very difficult to differentiate which IMFs are associated with crackles and which are associated with lung sounds. There are several approaches one can take to achieve this separation. In the literature, although not applied to lung sounds, researchers have used partial signal reconstruction based on the power threshold on IMFs [20], correlation comparison of IMFs and the original signal [21], power comparison [22], and using a fractal

dimension filter [23] to name a few techniques. Based on crackle simulation signals done by Karahiannis et al. [24] and applying EMD decomposition they found that fine crackles seem to start at IMF 2 and course crackle at IMF 3 but the ending IMF is unknown. Since this technique is faster than the other aforementioned technique, it will be used in this work.

2.1.4 Sparse Representation

Not a great deal of research has been done on the application of sparse based analysis on lung sounds. Current lung sound analysis techniques use frequency spectra, waveform and wavelet coefficients, which are not desirable since most of the time these features are not discriminative. The reason is that lung sounds as well as adventitious sounds overlap in both time and frequency domains [25]. Therefore, adaptive sparse representation is a better tool to use, because unlike the previous techniques, it is able to discriminate the adventitious sounds from the lung sounds. Separating them is essential for not only classification, but also for auditory display. Having only lung sounds and pure crackles enables a more accurate manipulation of sound parameters in sonification as well as improves discriminative sound enhancements in audification.

The theory behind sparse representation is the fact that signals can be efficiently represented by a small number of suitable basis functions. In order to do this, three assumptions must be made. The first assumption implies that the signal can be represented by a linear combination of basis functions plus a noise component (see equation below).

$$y = \sum_k y_k + e$$

The second assumption states that the signal can be a d-dimension vector containing d samples which is defined as y_k. If there exists an A_k with a d x n_k matrix whose columns are the vectors of d samples of the basis function, and an x_k vector with a dimension of n_k, then y_k can be represented as the matrix multiplication of the x_k which contains the coefficient for y_k expanded on the matrix A_k (see equation below).

$$y = \sum_k A_k x_k + e$$

The third and final assumption states that a small number of basis functions in A_k can synthesize y_k, mainly x contains a small number of nonzero scalar components or a sparse representation of the signal (see equation below). The sparse representation x can be obtained by solving a convex minimization problem and the l^1 regularization problem.

$$y = Ax + e$$

Based on prior knowledge of lung sounds, having a spectrum confined under 500Hz and adventitious sounds having a wide range of frequency components due to their explosive natures, a sinusoidal basis and a wavelet basis of the vesicular sound and adventitious sounds were used to find a sparse representation of each sound. Therefore the A_k matrix is comprised of a discrete cosine transform matrix and a Daubechies wavelet transform matrix for the lung sounds and adventitious sound respectively (see equation below).

$$y = [A_{DCT} A_{WT}] \begin{bmatrix} X_{DCT} \\ X_{WT} \end{bmatrix} + e$$

By solving the convex minimization problem and the l^1 regularization problem, a sparse solution can be obtained to recover the lung sounds and the crackles sounds using the equations below.

$$y_{lungsound} = [A_{DCT}]\,[x_{DCT}]$$

$$y_{advantitioussound} = [A_{DCT}]\,[x_{DCT}]$$

There are only two articles that explore the application of sparse representation of lung sounds by Sakai et al. ([26] and [25]) Sparse representation can also have a denoising property as well as can separate lung sounds from adventitious sounds effectively. This is advantageous to both sonification and audification as pure lung sounds and pure adventitious sounds can be more effective in the modification of parameters in sonification, and clean lung and adventitious sounds can be heard in audification which is further discussed in the next section.

2.1.5 Auditory Display

Auditory display (AD) essentially tries to use sounds in everyday life as Human to Computer Interface (HCI). The main goal of AD is to help us monitor and comprehend what the generated sound represents. In this work, we will focus on the non-speech medium of sounds which will try to exploit our evolutionary acquired adaptation to our environment which includes cognitive and preattentive cues, to try and understand the meaning of the AD generated sound. This work will mainly focus on two types of AD which differ in the amount of change and manipulation to which the original data is used to create the final output sound: audification and sonification. The main objective of this work is to identify which type of AD will be more suitable to discriminate normal and abnormal lung sounds.

2.1.5.1 Audification

Audification is the process of using the data as a direct translation of sound pressure values to the audible domain. In many cases, the data needs to be shifted into the audible domain and converted to

analog which will then be amplified so that humans can hear the sound generated from the raw data. Other than filtration of the generated sounds, no sound generating element is introduced. This means that the sound that is being generated has very minimal separation to the raw representation of the data.

In the case of lung sounds, they are already in the audible domain so little manipulation is needed to listen to them. On the other hand, enhancement to the adventitious portion of the lung sound can greatly influence our ability to discriminate abnormal from normal lung sounds. After the isolation of the crackle portion of the lung sounds (y_c) using the three signal processing techniques described earlier, there are two main audification techniques that will be explored;

1. Direct playing of the separated crackle sound.

2. Using a phase vocoder to stretch or compress the sound in time or shift the frequency of the sound in the frequency domain and then play it.

3. Frequency shift the signal to a higher frequency range.

2.1.5.1.1 Time Stretching

Due to the transient nature of the crackle sounds the crackling sound has a very short duration and can be easily missed by the physician who may wrongfully diagnose the patient. The goal of time stretching is to lengthen the sound with S samples to a new sound with $S' = \alpha * S$ samples with being the scaling factor. By lengthening the crackle sound we can identify the crackles more easily than playing them in normal speed. The challenging part of this is that one cannot simply stretch the perceived time without affecting the perceived acoustic properties of the signal such as pitch. For example, if $\alpha = 2$, one cannot simply stretch the signal by duplicating the samples because the frequency components will be compressed, and will result in the sound having a deeper tone. This is because a stretch in time will cause a compression in the frequency domain. Therefore an algorithm that performs time stretching must stretch the time but retain the frequency component.

2.1.5.1.2 Pitch Shifting

Also known as transposition in the music industry, it is the process of shifting the frequency component of the sound up or down in the frequency domain, thereby creating a higher pitched sound or a lower pitched sound. Similar to time stretching, one cannot simply vary the perceived pitch without modifying the speed/duration and the timbre (musical quality of the note which enables someone to differentiate one note from one another) of the signal. In signal processing, by stretching or compressing the spectrum, the pitch of the sound will be transposed up or down while the timbre is constant. Therefore an algorithm

which compresses or stretches the spectrum while retaining the speed or duration is needed to achieve pitch shifting. One possible solution is to do time stretching and re-sampling.

2.1.5.1.3 Phase Vocoder

Discovered by Flanagan and Goldern in 1966, the Phase Vocoder (PV) is a well-known signal processing technique that uses frequency domain transformations to create many digital audio effects such as time stretching, pitch shifting and spectral image processing. For the purpose of this work, we will focus on time stretching and pitch shifting. The algorithm works in three stages: the analysis phase, processing, and the re-synthesis phase.

In the analysis phase, the TF representation of the signal is obtained by calculating the Short Time Fourier transform (STFT). This is achieved by multiplying the signal $x(n)$ with a sliding window of length N, and then performing the Fast Fourier Transform(FFT) to obtain the magnitude and phase components of the signal. STFT has a poor frequency resolution without overlying harming time resolution artifacts. Therefore, in the processing stage, the PV uses phase information from the STFT to make improved frequency estimation.

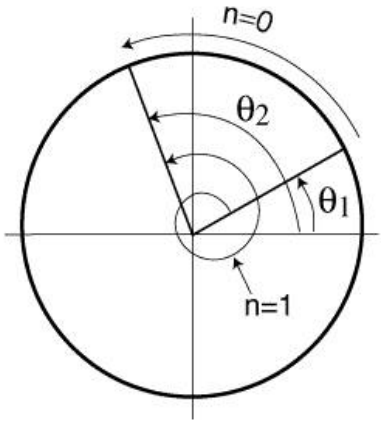

Figure 2.5: Unit circle example

Suppose there is a sinusoid of unknown frequency but with known phases. At time t1 the sinusoid has phase θ_1 and at time t_2 it has phase θ (see Figure 2.5). Looking at the unit circle there are two possibilities of movement, the first being that the sinusoid may have a frequency that moves it directly from θ_1 to θ_2 in time $t_2 - t_1$. Or secondly, it may begin at θ_1, move completely around the circle n times, and end at θ_2 after n full revolutions. Based on these two observations the frequency multiplied by the change in time must equal the change in angle (see equation below) and some multiple of 2π.

$$f = (\theta_1 - \theta_2 + 2\pi n) / 2\pi (t_2 - t_1)$$

Finally , in the re-synthesis stage, an inverse STFT (achieved by performing the inverse FFT and performing an overlap and then adding on the output of the inverse FFT) is performed to obtain the new modified signal. By varying the ratios between the window length and hop size (number of samples between each window) of the synthesis and re-synthesis windowing of the STFT and ISTFT, desired spectral characteristics can be achieved for different digital audio effects.

To achieve time stretching, the ratio between the hop size of the analysis (Ra) and Re-synthesis (Rs) reflect the ratio of the time stretched signal. For example, if the time ratio between Ra/Rs is 2 then the signal will be time stretched by a factor of two. The previous section stated that a possible pitch shifting technique is to increase time and re-sample. The PV achieves this by multiplying each grain of the STFT with the time stretching ratio Ns/Na where N represents the samples for each window of the analysis and re-synthesis stages, and performing a shift and add procedure with a factor of Na/Ns. This process can be seen in the diagram below.

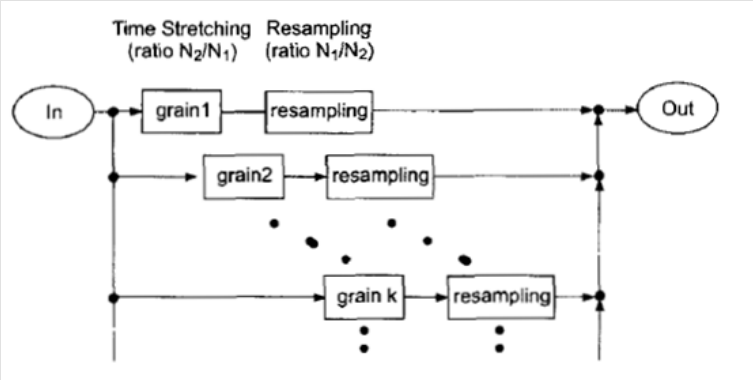

Figure 2.6: Block diagram of pitch shifting in PV taken from the DAFX: Digital Audio Effects on page 279.

2.1.5.1.4 Frequency Shifting

As described earlier, breath sounds have a wide range of frequencies, but many are near the lower threshold of human hearing. Therefore, it is sometime necessary for the sounds to be shifted higher in the frequency spectrum, in order for humans to hear them. This will be achieved by exploiting the frequency shifting property of the Discrete Fourier Transformation (DFT). The DFT property states that multiplying the signal with an Eigen function will result in a shift by w_o described in the equation below.

$$F^{-1}\left[X\left(jw \pm w_o\right)\right] = x\left[n\right] e^{\mp y w_o n}$$

2.1.5.2 Sonification

Sonification is the process of using the data to control a sound generator to monitor and analyze the data. Acoustic parameters such as pitch, onset, and envelope of the sounds from these sound generators are manipulated by extracting features through some type of mapping from the raw data. Selection of these features and the type of mapping function are crucial to take full advantage of the human ears' ability to detect abnormal sounds and outliers in the generated sound. By introducing a mapping function, complex structures of the data may be easier to comprehend as different data inputs cause one or more auditory variables to change. This work will explore three different types of sonification methods which will exploit the TF representation characteristics of the data. These characteristics will then be mapped to the MIDI paradigm where piano notes in a piano roll will be modified based on the TF map. The three methods are:

1. EMD and Hilbert Huang Spectrum (EMD-HHS) Sonification
2. Wavelet and Bump Source (WBS) Sonification
3. TF map of signal, based on Matching Pursuit (MP) Sonification

2.1.5.2.1 MIDI

Musical Instrument Digital Interface or MIDI as it is commonly known, is used in this work because of its real time capabilities. MIDI uses structured messages to control instrumental data from the computer, to many devices that support MIDI such as keyboards, synthesizers or even cell phones. These controls include channel and track numbers, notes to be played, velocity of the note, and duration to name a few. This work will focus on the modification of the MIDI encoding of pitch (see Table 2.1), note velocity, and the duration and timbre of the notes.

The pitch will be modified based on the frequency characteristics of the TF map, because high frequency components and outliers in the TF will have a higher pitch than others. Since crackles have a transient high frequency characteristic, the aim is to map these crackles to a higher frequency for easier detection.

The velocity of a note corresponds directly to the dynamic characteristics but not to the volume levels of the note. In the music world, these characteristics refer to sudden or gradual changes or relative loudness of the note and correspond to piano (soft), forte (loud), mezzo-piano (moderately soft), mezzo-forte (moderately loud), pianissimo (very soft), fortissimo (very loud). Velocity will be controlled by the energy component of the TF map resulting in an easier interpretation of the sound. For example, if the energy is high, then the note velocity is high, meaning a very aggressive note being played.

Finally, duration and timbre will be controlled by the duration of a particular frequency in time,

Table 2.1: MIDI encoding of pitch range from 0-127, encapsulating 10 octaves and 12 notes for each octave. For example the MIDI pitch encoded note for middle C (note C of the 4th octave) is 48.

	Notes											
Octave Number	**C**	**C#**	**D**	**D#**	**E**	**F**	**F#**	**G**	**G#**	**A**	**A#**	**B**
0	0	1	2	3	4	5	6	7	8	9	10	11
1	12	13	14	15	16	17	18	19	20	21	22	23
2	24	25	26	27	28	29	30	31	32	33	34	35
3	36	37	38	39	40	41	42	43	44	45	46	47
4	48	49	50	51	52	53	54	55	56	57	58	59
5	60	61	62	63	64	65	66	67	68	69	70	71
6	72	73	74	75	76	77	78	79	80	81	82	83
7	84	85	86	87	88	89	90	91	92	93	94	95
8	96	97	98	99	100	101	102	103	104	105	106	107
9	108	109	110	111	112	113	114	115	116	117	118	119
10	120	121	122	123	124	125	126	127				

based on the TF map. The timbre is affected by the envelope of the ADSR (see Figure 2.7), which stands for Attack, Decay, Sustain and Release of a note. To effectively explain the ADSR envelope, the process of playing a note on a piano can be examined as an example. The attack and decay of a note is essentially the initial push of the piano key, the sustained portion is the length in time in which the key is pressed and the release is the time at which the piano key is released. Within the MIDI paradigm this process is represented by the following equation, in which the modulated MIDI note is multiplied by the ADSR envelope, where n is a scaling factor to modify the relative loudness of the note.

$$note = ADSR_{envelope} * sin(2\pi * f_{note} * t + ADSR_{envelope} * n * sin(2\pi * f_{note} * t))$$

The duration and timbre are modified to manipulate the duration of the note being played. Essentially, the data is time stretched to identify the transient crackles. For example, if a crackle occurs in 1/32nd of a second, the human ear may not be able to detect it. But if the note is played which corresponds to crackle activity that lasts longer, the perceived length of the crackle sound is lengthened, thereby allowing for easier identification of the crackles.

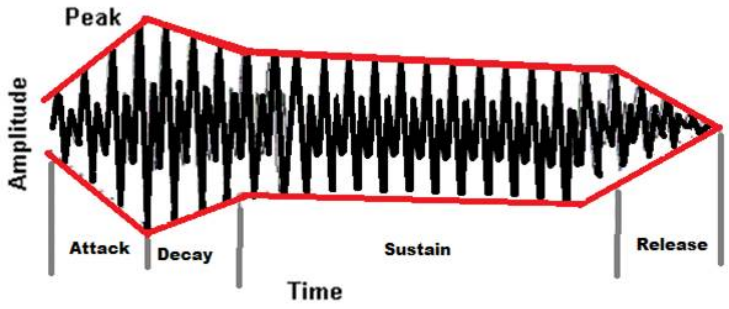

Figure 2.7: The envelope of an ADSR of a note.

2.1.5.2.2 EMD and Hilbert Huang Spectrum (EMD-HHS) Sonification

Described in earlier sections, EMD is data driven in that the basis of the expansion is of an adaptive nature which is confined by the data. This means it decomposes the data based on the patterns of the data itself, rather than a predefined basis function. This is highly attractive in sonification as patterns may arise as unique chords played when the data is sonified. Therefore a particular chord may correspond to abnormal behaviour while others correspond to normal activity of the lung sounds. To use EMD in our MIDI modification setup we must obtain energy, as well as TF characteristics of the IMF. To attain the TF representation, the Hilbert Huang Transform (HHT) is performed on each IMF, obtaining the Hilbert Spectrum (HS). The energy is acquired by squaring the absolute values of the HHT of each IMF. Since the spectrum of the lung sound goes beyond the number of MIDI encoded pitch values, the frequency is quantized by a predefined note range. For example, if one wanted to represent the HS in the 5th-6th octave range, then the TF will be quantized from 60-83 (based on Table 2.1). The flow chart diagram can be seen in Figure 2.8 below.

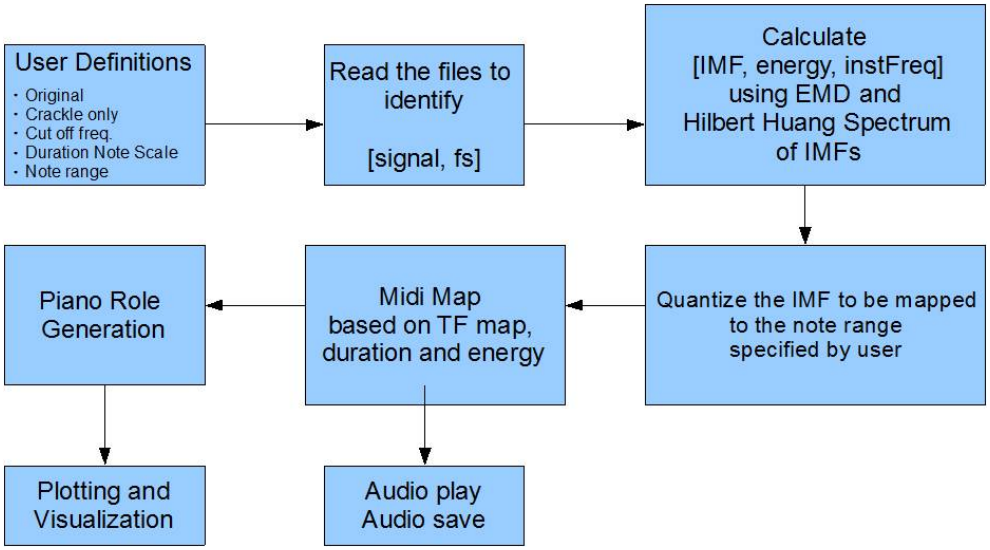

Figure 2.8: Flowchart diagram of the EMD-HHS sonification.

2.1.5.2.3 Wavelet and Bump Source (WBS) Sonification

As described earlier, due to the explosive nature of crackles, they have a transient characteristic in the time domain and broadband energy on the frequency representation. Wavelet is used based on the multi-resolution decomposition as well as its simultaneous analysis of the time and frequency domain. A direct mapping to MIDI of the TF map of the wavelet will result in a very noisy signal, due to the constant background respiration sounds. Rutkowski et al. [27] used both z-score and bump source modeling to achieve a better suited TF map to be used in the mapping onto MIDI. The main idea is to set predefined elementary parameterized functions called bumps whereby the TF map is represented by a set of these bumps. Bumps are half-ellipsoid functions defined by the equations below where u_f and u_t are the coordinates of the center of the ellipsoid and are defined by the t and f of the time and frequency index of the TF map.

$$\phi_b(f,t) = \begin{cases} a\sqrt{1-v} for 0 \leq v \leq 1 \\ 0 \ for \ v > 1 \end{cases}$$

$$v = \left(\frac{f-u_f}{l_f}\right)^2 + \left(\frac{t-u_t}{l_t}\right)^2$$

Before performing the bump algorithm on the wavelet TF map it is important to perform the z-score over the entire wavelet TF map. This is to provide high normalized amplitudes, and their location, which correspond to significant patterns of interest we want to sonify. This is because z-scores identify and describe the exact location of every point in a distribution. In the end of the bump modelling procedure

of the TF map, interesting patterns of the activity will be more accentuated from the background, which will be enhanced and played in the mapping to MIDI stage. The flowchart of the procedure can be seen in Figure 2.9 below.

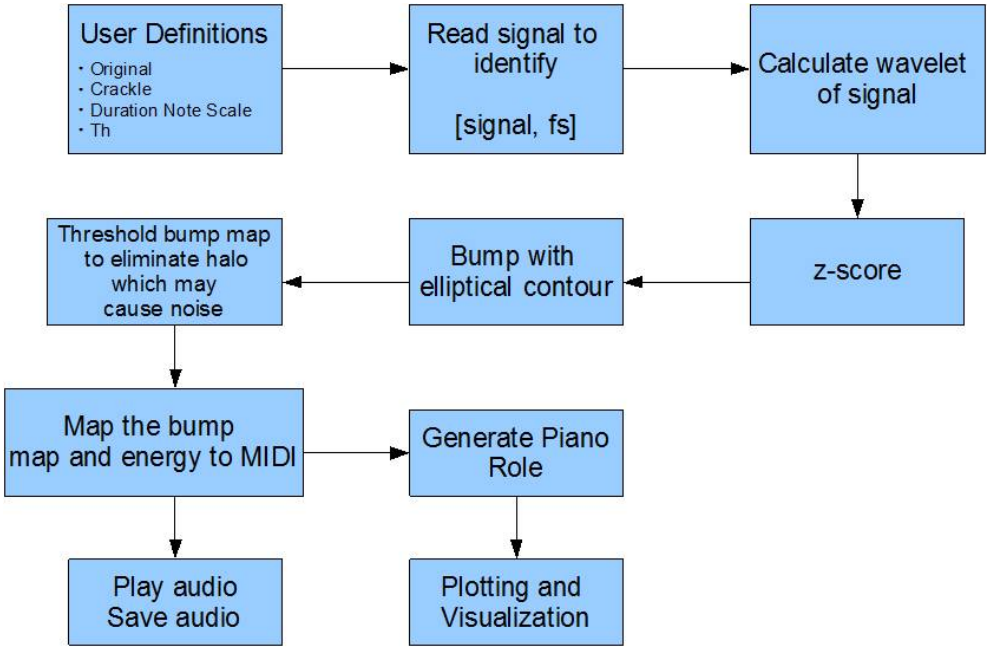

Figure 2.9: Flowchart diagram of the WBS sonification.

2.1.5.2.4 TF map of signal based on Matching Pursuit (MP) Sonification

The general approach to the MP algorithm [28] is to identify the best projection of a given signal $x(t)$ to an over-complete dictionary D in which the signal can be represented by the sum of weighted functions called atoms. Dictionaries, defined as D, are matrices that are real column vectors of length N and K number of atoms. Over complete dictionary is a collection of atoms in which the number of atoms exceeds the dimensions of the signal. In this way the signal $\vec{x}_a$ can be represented by linear combinations of atoms and sparse coefficient matrix $\vec{w}$, seen in the equation below.

$$\vec{x}_a = D\vec{w}$$

Another way to express $\vec{x}_a$ is as follows. Suppose that $\vec{x}_a$ is size $Nx1$, then it can be approximated by a linear combination of dictionary atoms and weights where the weights are sparse.

$$\vec{x}_a = w_1 d_1 + w_2 d_2 + w_3 d_3 ... w_n d_n$$

Since $\vec{x}_a$ is only an approximation of the signal, there exists an error r between the true signal and the approximated signal.

$$r = \vec{x}_a + \vec{x}$$

To obtain the minimum error, the calculations to obtain the sparse coefficient matrix must be minimized, as seen in the equation below

$$\vec{w}_{opt} = argmin_w \|w\| + \gamma \|x - Dw\|^2$$

In general, to obtain these weighted atoms, the MP algorithm identifies the atom which contains the largest inner product with the signal. The contribution of the atom is then subtracted and the MP algorithm then identifies the new largest inner product with the signal based on the updated signal. This procedure is repeated until the signal is decomposed or a particular iteration is reached. The idea of using MP is that it may provide a better resolution of the signal in the TF map due to the Gabor function, which is employed in this work to generate the dictionary. It will also provide a sparser signal which is more ideal when mapping the signal to MIDI for an audibly less noisy transformed sonified signal.

2.2 Signal Processing Techniques for Environmental Sound Analysis

As stated in the previous chapter, the ESA framework will be used to try and automatically detect inhaler sounds from the environment and to develop a program to quantitatively measure adherence to medication. Unlike speech processing, ESA is a branch of audio processing that tries to identify certain natural and artificial sounds in the environment and to automatically identify certain salient events in a recording. In this work, inhaler sounds are the environmental sound of interest. Most ESA technology has been applied for security purposes such as the detection of gun sounds[29]; in animal call detection of birds[30], and whales[31]; and also in applications to smart office[32] and smart homes for the assistance of the elderly [33] .

Because it is a relatively new idea in the world of audio signal processing, current ESA algorithms follow speech and music recognition theory. Nevertheless, many of the fundamental theories behind speech and music recognition theories cannot be directly applied to ESA due to the nature of the characteristics of environmental sounds. Unlike speech or musical sounds, environmental sounds have noise-like characteristics and have no structure. On the other hand, speech and music processing exploits our understanding of the inherent building blocks such as formants and phonetic structures[34] as well as musical structures such as melody and rhythm [35] to develop recognition algorithms. This deep understanding of their structures allows us to use modeling techniques such as the Hidden Markov

Model (HMM). Therefore, many of the features that are extracted must exploit the non-stationary characteristics of environmental sounds. This section will look at these non-stationary features as well.

2.2.1 Signal Segmentation

Inhaler sounds are non-stationary, meaning statistically they change over time. To address their non-stationary property we segment the signals in short epoch or frames in which we can assume that in that short amount of time they are stationary. In the world of speech and environmental sound analysis there are three such segmentation techniques: frame based, sub-frame based and sequential based.

In frame-based segmentation the signal is segmented in epochs or frames that are "x" seconds, which are overlapped by 50 percent. In each frame features are extracted. The problem with this method is that if the frame is too short, the long term characteristics of the signal are lost and if the frame is too long then the inter event characteristics are lost.

In Sub-Frame-Based Segmentation, after segmenting the data based on the frame-based method, the frame is further divided into frames usually 25msec with an overlap of 10msec. This more granular division is able to extract non-stationary features as well as model intra frame characteristics when used with a sequential classifier such as a Hidden Markov Model (HMM). The main drawback of this method is the resultant feature matrix is large but feature reducing algorithms such as Principal Component Analysis (PCA) and averaging can reduce the feature matrix for a more robust feature set.

In sequential based segmentation the signals are divided into smaller windows as compared to frame-based processing and each segment is classified as an important building block for an event. This is usually used in speech processing where each frame is labeled as a phoneme and the event is the word. Based on these three methods chosen the sub-frame based method was chosen.

2.2.2 Feature Extraction

As stated earlier, before extracting features, it is important to realize that environmental and inhaler sounds have no sound structure and are essentially noise like. Unlike speech processing where the events (words) have building blocks (phonemes), inhaler sounds have no such foundations. If heard at short epochs, the sound characteristics are transients that have no associated meaning such as melody or rhythm. Keeping this in mind, the following features were chosen to represent the noise-like properties of the inhaler sounds. The properties include their transient and chaotic nature.

2.2.2.1 Mel Frequency Cepstral Coefficients (MFCC)

In the world of speech and music recognition, the most widely used feature is the Mel Freqeuency Cepstral Coefficient (MFCC). Introduce by Davis and Mermelstein in the 1980s [36], this algorithm models our understanding of the inner workings of the human ear, described in Chapter 1. The implementation of the algorithm is as follows;

1. Segment the signal into short epoch/frames.
2. For each epoch/frame calculate the power spectrum (these two steps are also known as Short Time Fourier Transform STFT).
3. Apply the mel filter bank and sum the energy in each filter.
4. Take the log of the energies.
5. Perform Discrete Cosine Transform(DCT) on the log energies.
6. Keep only 1-n numbers of MFCCs you need.

The first step is performed to tackle the non-stationary behaviour of the signal by segmenting it into short epochs called frames, which was explored in detail in the previous section. In the second step, the purpose of obtaining the power spectrum is purely motivated by the human cochlea. As sounds enter our ear, the cochlea vibrates at different areas and at different strengths. The vibrations cause tiny hairs to also vibrate and nerves that are attached to them send frequency information to the brain. In short, these hairs act like the frequency bins in the STFT.

Humans cannot discern frequencies that are closely spaced and are more exaggerated at higher frequencies. This means that we perceive small changes of pitch better at low frequencies than at high frequencies. The perception of these frequencies by our brains is modeled though the use of the Mel filter bank. The characteristics of the filters of the Mel filter bank are narrow at lower frequencies and get wider at higher frequencies. The energy is summed in each filter to get a rough idea of how much energy exists in different frequencies.

The logarithm of these energies is taken because humans do not linearly hear loudness. In general, eight times as much energy is needed for us to perceive a doubling in volume. Therefore, if there are large variations in energy, the sound perceived may not be the same because the volume of the sound may already be high. The logarithm is used instead of a root function because the logarithm enables us to use channel normalisation techniques such as cepstral mean subtraction. Due to the overlapping nature of the mel filter bank, the energies are correlated with one another. So DCT of the log filter bank

energies is taken to de-correlate the energies. In this work, thirteen coefficients are kept because higher coefficients represent fast changes in the energies which may degrade classifier performance.

In many works, the delta and delta coefficients are also calculated to determine the rate of change and how fast these changes occur. They are essentially the first and second derivatives of the MFCCs and can be thought of as speed and acceleration of the changes. The delta coefficients are calculated using the MFCC's and the following equation,

$$delta_i = \frac{(\sum_{i=1}^{Numberof MFCC} n(c_{t+n} - c_{t-n}))}{2\sum_{i=1}^{Numberof MFCC} n^2}$$

The delta-delta coefficients are obtained using the same equation but using the delta coefficients instead of the MFCC coefficients. These delta and delta-delta coefficients can be thought of as the speed and acceleration of change of the MFCC's.

2.2.2.2 Gammatone Frequency Cepstral Coefficients (GFCC)

Similar to the MFCC, the GFCC uses the gammatone filter bank to model human hearing instead of the Mel filter bank. While the Mel filter bank is modeled from the behaviour of the cochlea, the gammatone filter bank is modeled by the motion of the basilar membrane, which is a thin membrane within the cochlea. The impulse response of each filter in the gammatone filter bank has a smother shape than that of the Mel filter bank which may capture the characteristics of the inhalers better.

The motivation of the use of GFCC in this work stems from the use of gammatones in the surveillance ESR work done in [37] to identify footsteps and gunshots. Valero and Alias showed that gammatones were able to characterize the transient nature of footsteps and gunshots. Their work also showed that the GFCC and the MFCC complemented each other, increasing the classification accuracy when combined. Therefore in this work, since the inhalers also have this transient behaviour, we will use the GFCC in conjunction with the MFCC.

2.2.2.3 Time and Frequency Based Features

In the beginning of this section we have established that environmental sounds lack fundamental building blocks and that we have no prior knowledge of their structure. Therefore, we must extract as many relevant features as possible to try and describe the sounds. In this section we will explore other features which may help in discerning the inhaler sound from the environment.

To quantify the variability of the volume, and the volume itself as a function of distance between the inhaler and the microphone, the short time energy and the root mean square of the signal was calculated. The distribution of the signal and quiet segments such as pauses can indicate quite environments and

calculating the spectral centroid and spectral roll off can improve classification. Due to the chaotic unstructured nature of the sounds, spectral entropy is also calculated. The mean and standard deviation of the autocorrelation function of the signal was calculated to obtain an impression of signal frequency characteristics.

Finally, a three layer discrete Myer wavelet decomposition of the signal was performed and mean and standard deviations of each layer was obtained. The incentive for the addition of the wavelet features was driven by the works of [37] and [38] where it was shown that wavelet features were on par with the MFCC features. Adding them to the pool of features may result in creating a complementary set of features which can improve classification accuracy.

2.2.3 Feature Selection

Earlier in this chapter it was established that this work will focus on the Sub-Frame-Based segmentation scheme. As a result, for each sound segment of half a second in duration, and sub-frames processed with 25 msec and an overlap of 10 msec, forty-eight frames are considered for every half second of recording. Furthermore, 97 feature elements are extracted based on the features that were discusses in the previous section. This results in a feature matrix with a dimension of 48x97 consisting of 4656 elements. This is too large for classification purposes. In general, the number of features should be equal to or less than the number of samples. In this case, we have twice as much features as samples so we need to apply some feature reduction techniques. This is important to prevent over-generalization, remove redundant features which may hinder classification, as well as identify complementary features. This work has employed the maximum relevance and minimum redundancy (mRMR) algorithm to do this task [30].

Essentially, the main goal is to simultaneously address relevance and redundancy. Features are relevant if and only if they are highly valued as features that are good in discriminating the classes (in this case environment and inhaler). Meanwhile, features that are redundant do not add any more meaningful information in discriminating classes. Therefore this may give rise to features that are independent of each other and complement each other in such a way that they discriminate the two classes.

Since different environmental regions may have different characteristics and may need different features to discriminate the environment from the inhaler sounds, the mRMR feature reduction algorithm must be performed for different regions. This process will result in obtaining different features as well as different feature lengths for each region. Intuitively, this makes sense as in quiet environments the algorithm may only need a couple of features to distinguish the inhaler sounds from the environment. While on the other hand, regions that have a lot of noise, may need more features to distinguish the

inhaler sounds.

To determine complementary features, different scenarios were implemented where each feature vector had a subset of the features described in the previous section. The list of feature scenarios are listed below:

1. MFCC and its delta derivatives combined with TF features
2. GFCC and its delta derivatives combined with TF features
3. MFCC and GFCC and their delta derivatives
4. All features combined

2.2.4 Classification Method Using Hidden Markov Model (HMM)

Without going into the detailed mathematical reasoning behind the HMM, the main objective of a HMM is to produce a model based on finite state machines and observed patterns. In our case, the observed patterns are the features described in the previous section associated with the inhaler sounds or the environment. Because we do not know the patterns of the spectrum, guessing is required which is done in the hidden layer of the HMM. Within the hidden layer, statistical parameters of the changes of the state machines are calculated using the iterative optimization problem (this work uses the Baum-Welch method for training and testing). Therefore, during the training stage, the model learns the patterns of the feature given as inputs. A graphical representation can be seen in the Figure 2.10 below.

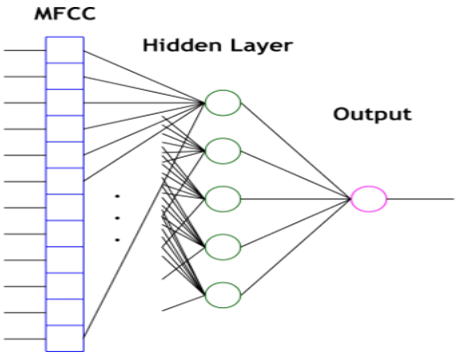

Figure 2.10: Illustration of steps for HMM.

The Hidden Markov Model can be thought of as a finite state machine in which each state changes depending on the initial conditions, probability distribution, and observed sequences. As users, we can only see the observed sequences (features) and we do not know the steps in which the features relate to each other to obtain an output, in this case inhaler or environment. There are three main things to think about when using the HMM: type of model, the number of states, and the number of layers.

There are two commonly used models: the ergodic and the left-right model [39]. In the ergodic model, every state is connected, while the left-right model has the property that as time increases the state index also increases. This work uses the left-right model as it can model signals that are changing over time.

The number of states relates to the number of fundamental sounds heard or the number of classes. There are two possibilities to consider, the first being using two states to represent the two classes (environmental and inhalers sounds). Meanwhile the second is using three states to represent the attack, sustain, and release of sounds described in the previous section. Though the previous section directly correlates ADSR to notes being played, the ADSR can be generalized to all sounds. Consider the sound of knocking. The first strike of the door creates the attack, albeit short. The duration of the knock is then followed, and finally the release is the dampening nature of the knocking sound. In contrast, the inhaler sound also has this attack, sustain and release nature, when the inhaler is pressed, it releases a hissing sound and then there is dampening of the hissing sound.

The last parameter to consider is the number of layers. Typically in speech processing, three layers are used: word/dictionary/event, model (phonemes), and hidden. In the case of this work, we do not need the model layer because as stated in the previous section, environmental sounds do not have fundamental building blocks such as phonemes. Therefore, the number of layers is confined to two. The first layer is the event layer, and the second layer is the hidden layer which is associated with the features of the events.

Hence, this work uses two HMM classification schemes, the first being a two state two layer left-right HMM and the second being three state two layer left-right HMM. In this work classification was done by selecting 50 percent as training data and 50 percent as testing data.

2.3 Noise

Noise is purely defined by the type of application and context. In some applications, certain sounds are thought of as noise while in others those same sounds can be thought of as signals. This work treats both crackles, and inhaler sounds, which are transient noise in the lung and environment, as signals. Both in terms of signals and sounds perspective, crackles and inhaler sounds are noise because of their transient non melodic nature and their structure less appearance.

In the real world, signals are often corrupted by noise and breathing sounds are no exception. Lung sounds have relatively low frequencies and low intensities, and it is essential to remove the noise and other ambient sounds from the recorded lung sounds before diagnosis. Some examples of these types of

noise include heart sounds, respiratory muscle sounds, friction rubs, and ambient noise. This noise affects our ability to obtain important information from the breathing sounds and can hinder our performance in diagnosing and treating different types of respiratory illnesses.

It is inevitable for lung sounds to be contaminated by heart sounds, as the lung and heart are in proximity to each other, and they are considered as two independent source signals. However, due to the nature of the sounds within the lungs such as reflection and delays, the mixed (lung and heart) recorded signals are not instantaneous but of a convolution mixture of the two signals. One solution is to use independent component analysis (ICA) to separate the lung sounds from the heart sounds. Others include using wavelet, short time Fourier transform and fuzzy logic.

In terms of the environmental sound analysis for inhaler detection, the environment is treated as noise and inhaler sounds as signal. To do noise and signal separation, a HMM classifier is trained to classify what is signal and what is noise based on labeled training samples. Segments of signal features and noise features are fed into the binary classifier and are labeled either as inhaler sounds or environmental sounds. The following chapter lays out the AD techniques performed in this work and the results obtained.

Chapter 3

Auditory Display of Respiratory Sounds

3.1 Motivation

As previously stated in Chapter 1, current auscultation techniques are confined to physicians either using digital or regular stethoscopes to listen in on breath sounds and to identify abnormalities such as wheezing or crackles. These abnormalities can be early symptoms of a more deeply rooted respiratory ailment. In many cases, performing auscultations must occur in quiet environments to properly detect these abnormalities. Furthermore, auscultation is purely subjective as studies such as [40] have shown that doctors cannot fully pick out the crackling noises of patients. This chapter will explore how AD, in the form of audification and sonification, can benefit from the enhancement of crackle detection in respiratory lung sounds via digital stethoscopes.

3.2 Methodology

In Chapter 2, two different types of AD were explained: sonification and audification. The mapping and techniques used for sonification and audification are directly affected by the type of application. Different applications need different types of mapping, and different applications perform better in the sonification context and some perform better under an audification context. In a sense, there is no real science behind it. Identifying the proper mapping and technique is purely done experimentally to identify the ideal method to represent the data, in such a way that the sound produced can exploit the

human auditory systems ability to discern patterns and outliers. Hence, due to this experimental nature in identifying the ideal AD, different sonification and audification techniques will be explored in this work. In addition, as stated in Chapter 2 due to the importance of crackle sounds and its diagnostic value in early detection of respiratory illnesses, the crackle sounds are extracted from the lung sounds and are used as signals to be audified and sonified.

The two-step proposed audification and sonification schemes can be seen in the flowcharts on Figure 3.1 and Figure 3.2. For both AD schemes, the first block separates the crackle sound (y_c) and the vesicular sounds (y_v) from the lung sounds(y) using either EMD or wavelet packet thresholding. Then, in the audification scheme the extracted crackle sounds can be either played, frequency shifted such that the crackle sound can be more easily heard, or the time and frequency can be manipulated by the phase vocoder to enhance the characteristics of the crackle sounds. Since different crackles can correspond to different diseases it is at this point that AD should highlight different types of crackles. In this work we will try and differentiate fine and coarse crackles. Note that all lung sounds were obtained through the R.A.L.E. database.

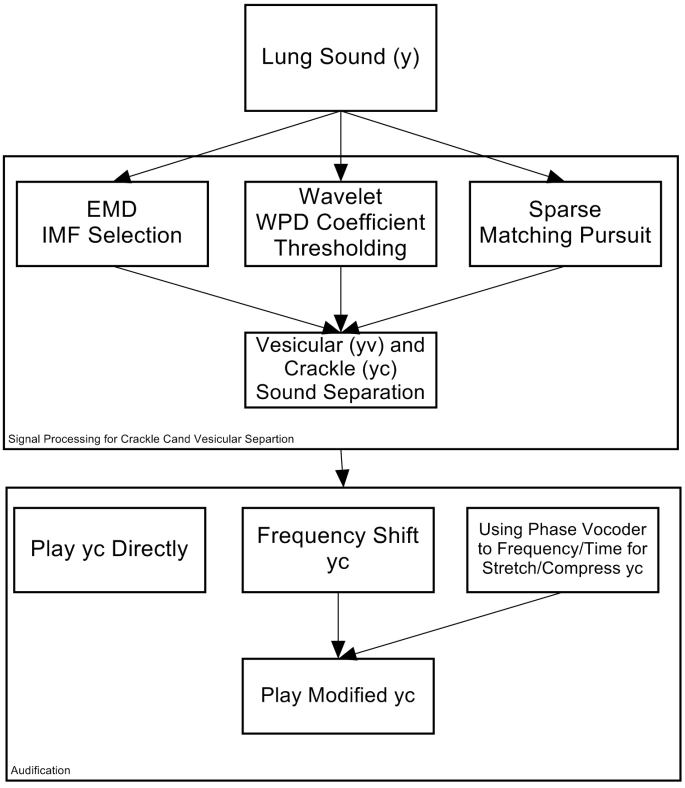

Figure 3.1: Audification of lung sounds block diagram.

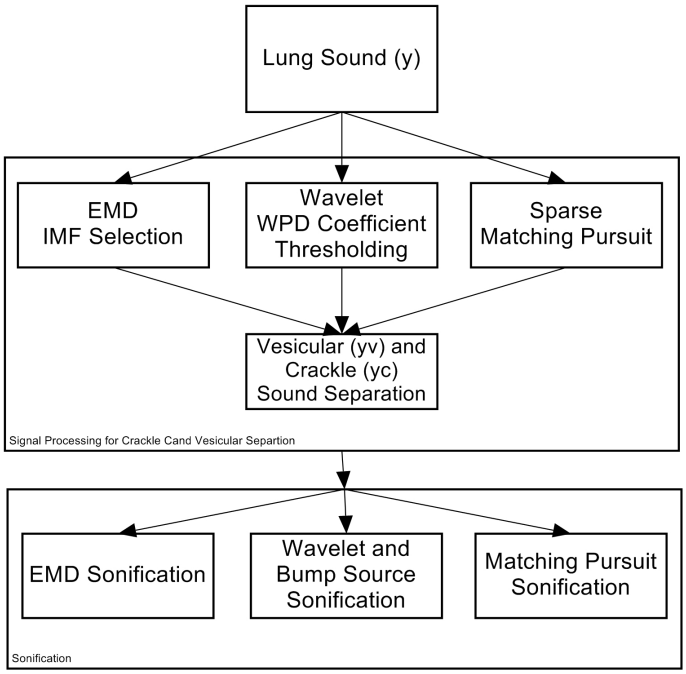

Figure 3.2: Sonification lung sound block diagram.

3.3 Crackle and Vesicular Sound Separation Results

As previously stated, before AD is performed, crackle and vesicular sounds must be separated and only the crackle sounds, which are hard to distinguish, will be subjected to AD. This section shows results from the separation of vesicular sounds and crackle sounds using EMD, wavelet thresholding and sparse separation. In the case of EMD, it was very difficult to identify the IMF's which solely corresponded with crackles and vesicular sounds due to mode mixing (seen in Figure 3.3 through Figure 3.5).

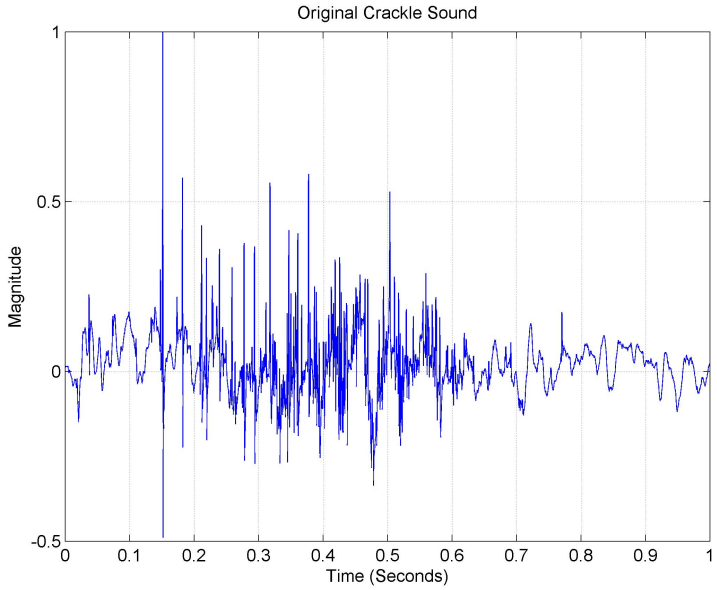

Figure 3.3: EMD decomposition of a lung sound containing crackles with 10 IMFs.

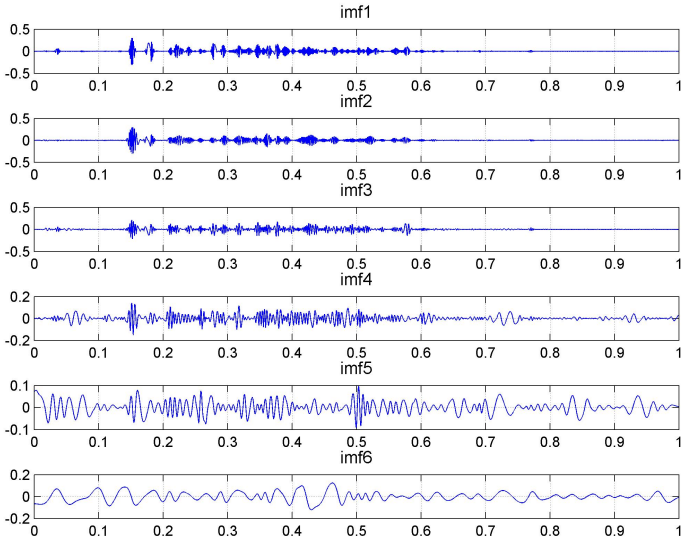

Figure 3.4: EMD decomposition of a lung sound containing crackles showing IMF 1 - IMF 6.

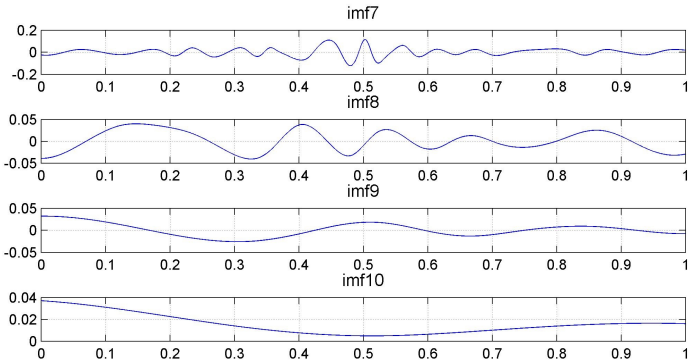

Figure 3.5: EMD decomposition of a lung sound containing crackles showing IMF 7 - IMF 10.

In the case of the sparsity algorithm, the separation of crackle sounds from vesicular sounds was not achieved, as the output of the algorithm was neither the sound of a crackle nor that of a vesicular sound on inspection of listening to the separated sounds. On the other hand, the wavelet thresholding results showed great separation of the crackle sounds and the results can be seen in Figure 3.6. The original lung sound can be seen in the upper part of the figure while the extracted crackle component can be seen in the lower part of the figure. Based on this result, this algorithm was successful in extracting crackle sounds from the original lung signal.

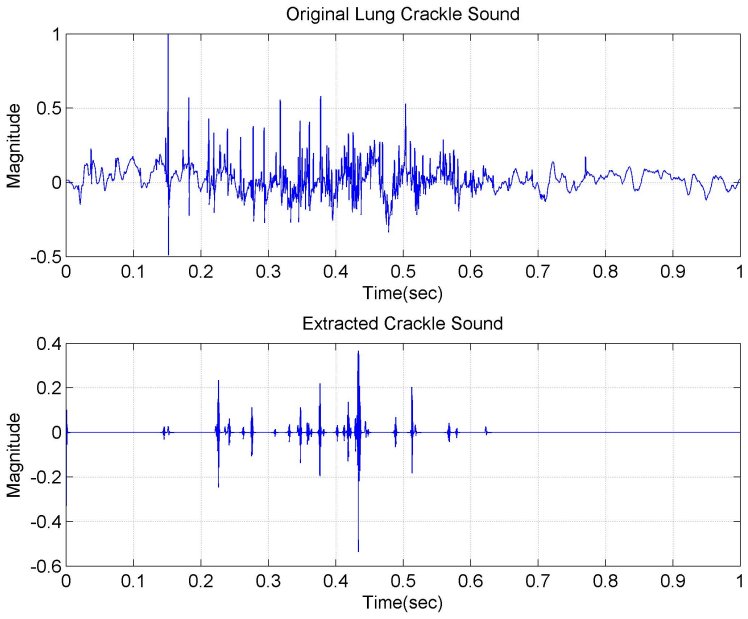

Figure 3.6: Result of applying wavelet algorithm to a one second lung recording.

Based on speed and accuracy, the wavelet approach was the best in separating the crackle sound and vesicular so its output will be used for AD.

3.4 Identifying Possible False Positive

It is important to apply the WPT separation algorithm to other well known normal and abnormal respiratory sounds to test for robustness. Robustness in this case means the ability of only identifying crackles in crackle adventitious lung sounds. The three figures below show the results of applying WPT to a strider and wheezing sounds (which are both abnormal) as well as to a vesicular lung sound (which is normal). It can be seen that the algorithm correctly did not identify any crackle sound components with in the wheezing and vesicular sounds. With respect to the stridor sound, the algorithm detected some minute low amplitude crackle sound components. Compared to the large amplitude transient spikes in the course crackle seen on Figure 3.6. It can be noted that this type of false positive can be ignored by simply applying a thresholding criteria to the algorithm.

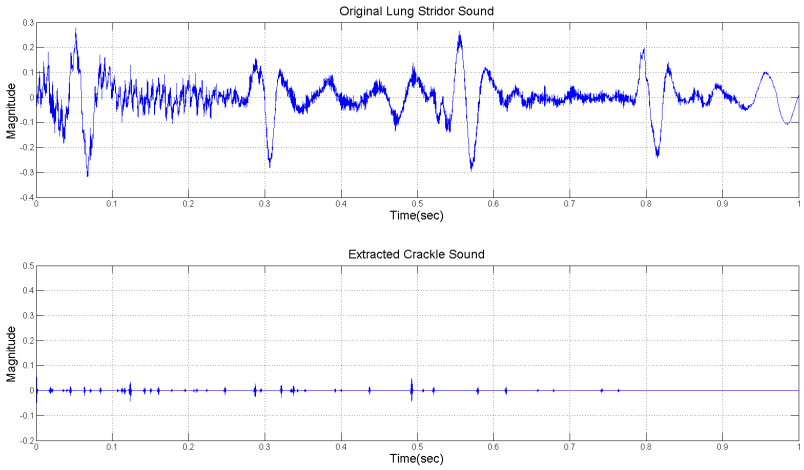

Figure 3.7: Application of WPT algorithm to Stridor.

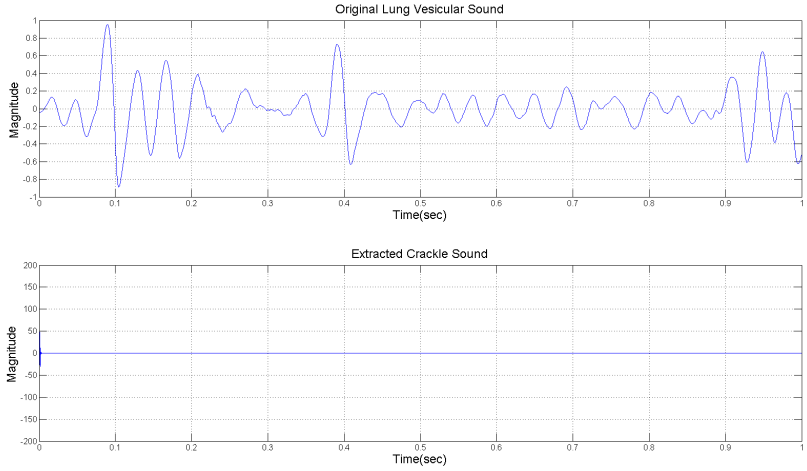

Figure 3.8: Application of WPT algorithm to Vesicular.

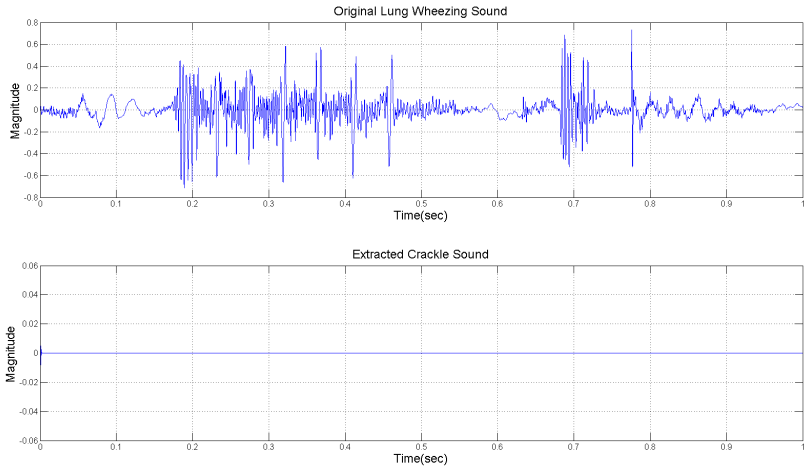

Figure 3.9: Application of WPT algorithm to Wheezing.

3.5 Audification Results

For the ease of testing multiple signals with various changing parameters for both the audification and sonification algorithms, a GUI (Graphical User Interface) was developed which can be seen in Figure 3.10.

The original files with both crackle and vesicular sounds can be loaded as well as the output of the wavelet algorithm (i.e. the separated crackle sound from the vesicular sound). After that, several parameters in each of the sonification and audification panels can be changed and the modified sounds can be played and compared with the original recording using the play sound panel.

In this work, two recordings will be presented which files are corresponding to fine crackles and coarse crackles which were explained in Chapter 1. The hope is that different types of crackles will be distinguishable from each other based on either their sonified or audified sounds. For each sound file, it is important to note that it is very difficult to distinguish between the fine or coarse crackle sounds based on gross visual inspection and the sounds that they produce.

All the audification results can be heard in the attached folder under the audification subdirectories of each folder pertaining to fine and coarse crackles or on the website [41]. Using the GUI, each file was frequency stretched or pitch shifted to slow down each crackle by a factor of 2, 4, and 8. Frequency stretching and pitch shifting created a very high pitched version of each crackle. Though audible, the fine and coarse crackles were still very difficult to differentiate from each other, even with the prolonged exposure of time. It was also still very difficult to differentiate the crackle sounds from each other even after shifting the signals to a higher frequency in the frequency spectrum. Based on these reasons, it

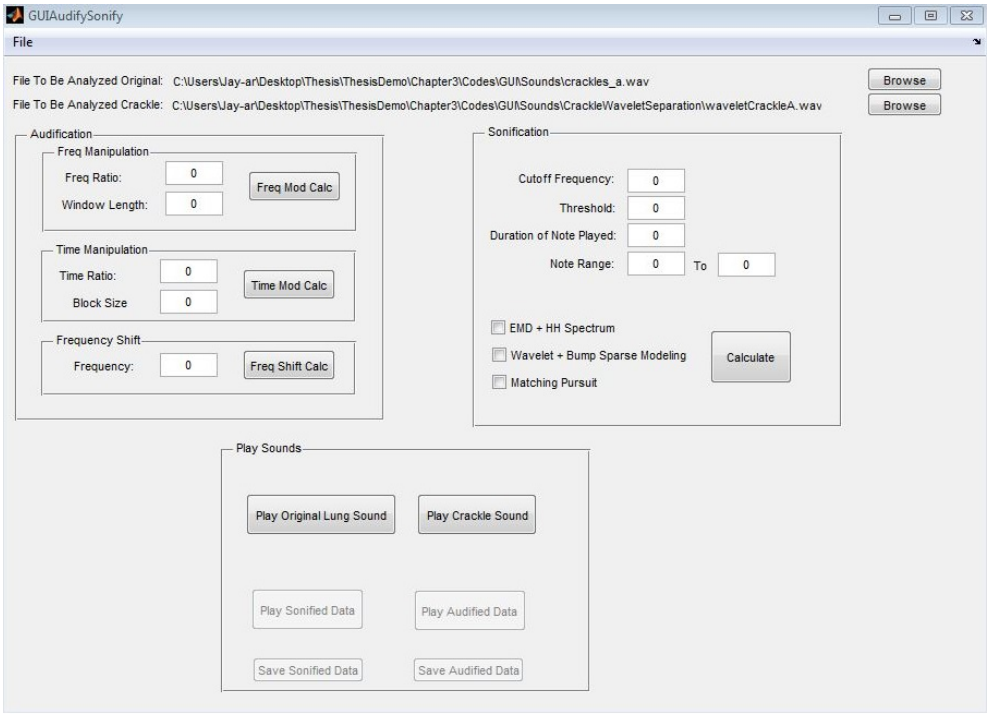

Figure 3.10: GUI for sonification and audification testing.

seems that audification is not a good AD to be used in distinguishing different crackles.

3.6 Sonification Results

The sonification mapping technique performed in this section was described in Chapter 2 and the GUI seen in Figure 3.10 was also used to easily apply these techniques to multiple sounds. To reiterate, the different TF maps (HHS,WBS and MP), were mapped according to the MIDI paradigm, where MIDI parameters, which were pitch, velocity and duration, were controlled via TF characteristics, energy and duration of frequency in time respectively. All the sonification results can be heard in the attached folder under the sonification subdirectories of each folder pertaining to fine and coarse crackles, while their corresponding graphs can be seen below.

3.6.1 Crackle A Results

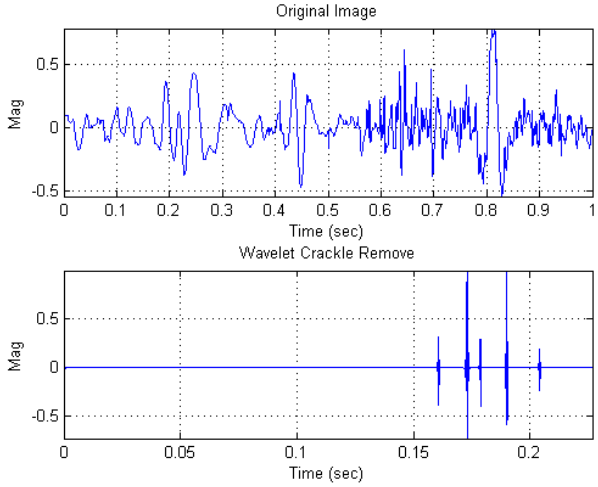

Figure 3.11: The top graph shows the waveform of the original vesicular and crackle sound while the bottom waveform shows the separated crackle sound from the vesicular sound based on the wavelet packet algorithm.

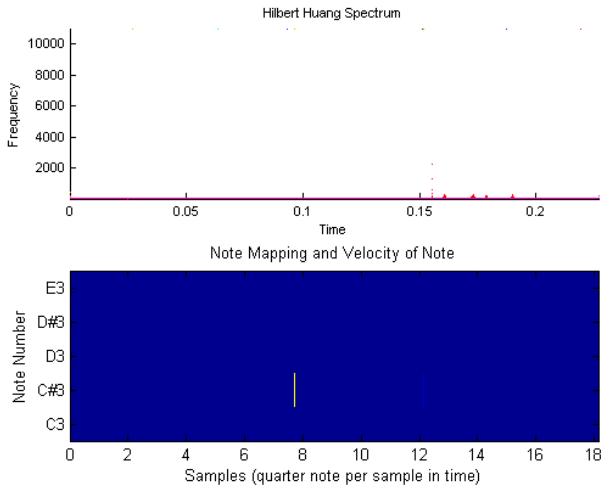

Figure 3.12: The top most figure shows the Hilbert Huang Spectrum, where each dot represents a frequency. The bottom plot shows the note mapping of the frequencies.

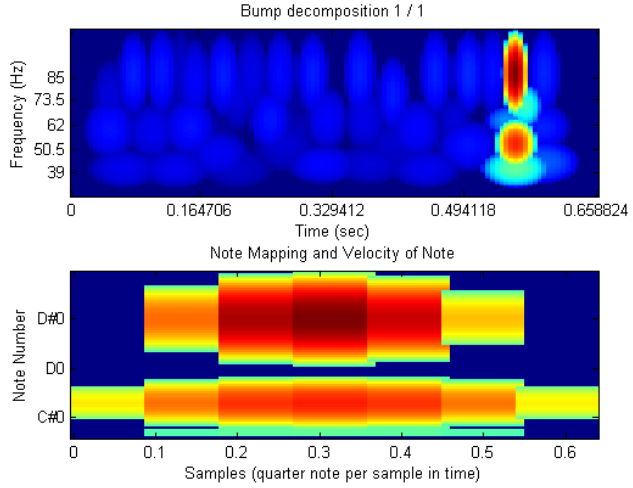

Figure 3.13: The top most graph is the TF map output of the WBS sonification algorithm while the bottom graph shows the corresponding MIDI map. The darker the note, the higher the note's timbre.

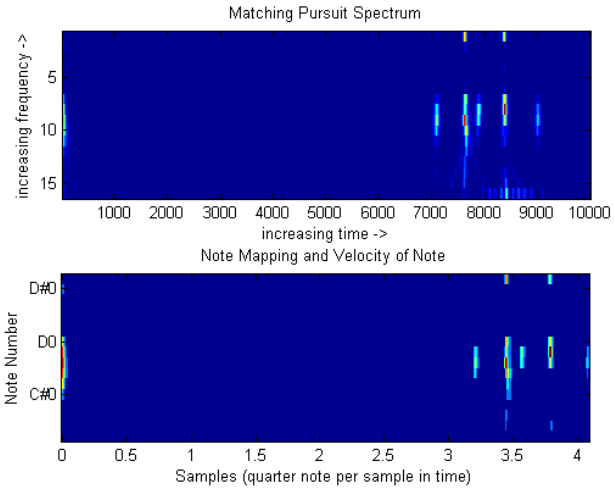

Figure 3.14: The top most graph is the TF map output of the MP sonification algorithm while the bottom graph shows the corresponding MIDI map. The darker the note, the higher the note's timbre.

3.6.2 Crackle B Results

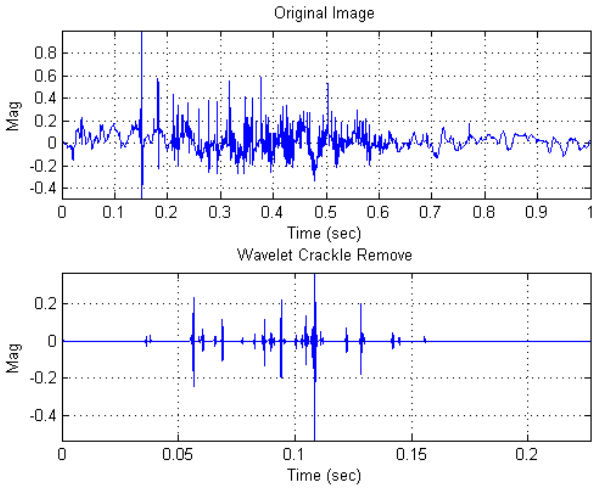

Figure 3.15: The top graph shows the waveform of the original vesicular and crackle sound while the bottom waveform shows the separated crackle sound from the vesicular sound based on the wavelet packet algorithm.

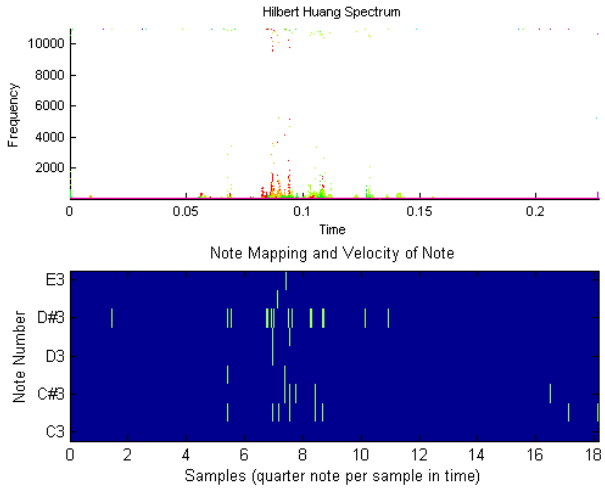

Figure 3.16: The top most figure shows the Hilbert Huang Spectrum, where each dot represents a frequency. The bottom plot shows the note mapping of the frequencies.

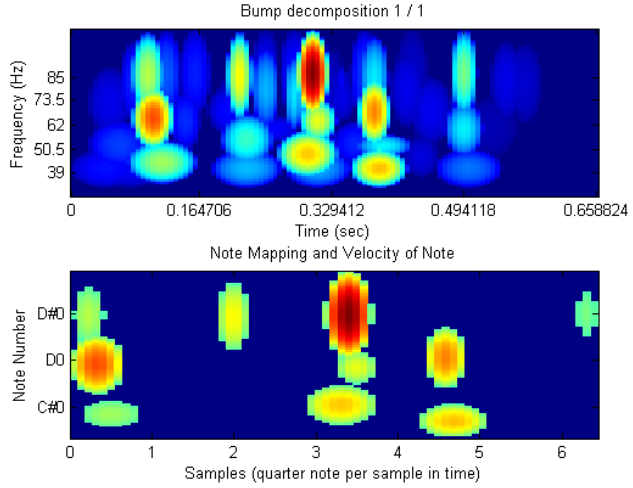

Figure 3.17: The top most graph is the TF map output of the WBS sonification algorithm while the bottom graph shows the corresponding MIDI map. The redder the note, the higher the note's timbre.

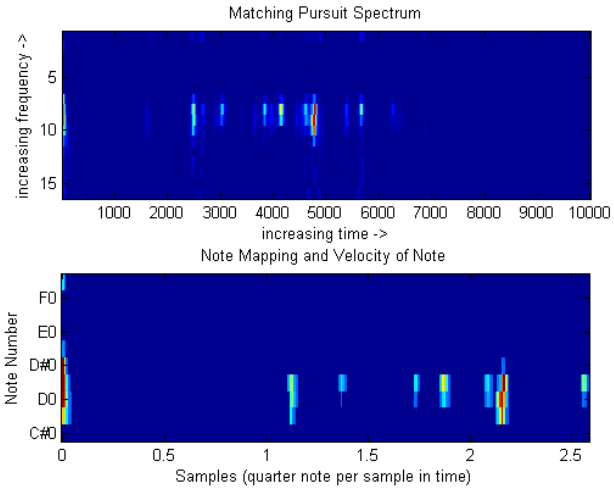

Figure 3.18: The top most graph is the TF map output of the MP sonification algorithm while the bottom graph shows the corresponding MIDI map. The redder the note, the higher the note's timbre.

3.7 Discussion

It is clear by visual inspection of the note mapping and listening to the audio files, that HHS, WBS and MP all produced very distinct audio sounds. In the case of EMD-HHS it was hypothesised that by using EMD and HHS, chords would appear which correspond to the decomposed signals. Based on Figure 3.12 and Figure 3.16, fine crackles did not produce chords, but crackle B did. This suggests that different crackles will produce different chords. Chords appear as multiple notes being played at a given time on the MIDI mapped TF. An important discovery is that different crackle bursts correspond to different chords. This may mean that to characterize different crackle types, the relationship between different chords must also be considered. Though this will add another layer of complexity, the human auditory system is acute to detect these patterns. This is apparent with experienced piano players who can replicate the song they just heard without seeing the score. This is because they can differentiate the different chords being played and they can associate them with piano keys. Similarly, if one were to master the different types of chords and how they relate to each other in terms of a musical piece, then it is possible for a person to get an impression of what type of crackle is being sonified.

In the case of WBS it was hypothesised that the wavelet can be used to give a multi-resolution and a TF analysis of the crackle sounds. Though the direct wavelet transformation would yield a very noisy sonification, the bump source sonification enhanced the most salient bumps. In the TF representation of both crackles in Figure 3.13 and Figure 3.17, light blue bumps can be clearly seen. These light bumps are the noisy portion of the signal that was brought out through the wavelet transform. But through

thresholding and bump source modeling, salient features of the TF maps were highlighted resulting in distinct notes being played. For fine crackles, two notes were played (C ♯ and D♯) simultaneously throughout the crackle. On the other hand, for crackle B, three chords consisting of C♯, C and D♯ were played along with 2 single instances of D♯ being played.

Finally, in the case of sparse sonification using MP,seen in Figure 3.14 and Figure 3.18, short bursts of sounds correspond to crackle bursts. The distribution of these bursts of sounds corresponds to different types of crackles. For fine crackles, the beeps are close together and come in fast successions, while on the other hand for crackle B, the beeps are spaced farther apart. The spacing of these beeps comes about through the sparse representation using MP.

Based on listening and looking at the spectrum as well as the note mapping graphs in each of the sonification methods, one can determine the inherent audible differences between the fine crackle and coarse crackle sounds. It is also clear that the short burst characteristics of the crackle sounds are enhanced and are easier to identify compared to the original lung recordings, where although crackles can be heard, the number of crackles are hard to determine. However, even if the sonified sounds produced by coarse and fine crackles are heard, it is still very difficult for the average person to identify which of the sonified sound is associated with a fine or coarse crackle sound, unless they have prior knowledge of the characteristics of the chords and notes being played in the sonified versions. Therefore, although sonification is a better representation to discriminate the crackle sounds, it cannot be easily integrated into the medical field today, as doctors have to be trained to use such new techniques. However, students who are becoming doctors can take advantage of this approach as it is easier to discriminate the crackle sounds for early diagnosis of abnormalities, because they do not currently have existing experience of listening to raw lung sounds. This can therefore be a training tool for aspiring doctors.

In the subsequent chapter, the ESA framework proposed by this work will be explored, as well as the results and their implications to its application of inhaler sound detection for adherence of asthma medication.

Chapter 4

Inhaler Detection Based on ESA Techniques

4.1 Motivation

ESA is a new field of audio signal processing in which salient or important sounds are identified in the environment. The features for the ESA system used in this work are the MFCC and GFCC and their first and second derivatives, along with a list of TF features mentioned in Chapter 2. To prevent over generalization caused by the large feature vector and to improve discrimination capabilities, the mRMR algorithm was used to reduce the feature space. Finally, 2 two layered HMMs , a three state model and a two state model which was mentioned in Chapter 2, were used to model the system.

As previously stated, asthma is a prevalent respiratory disease and the mitigation of effects and symptoms are difficult due to the lack of medication adherence from the patients. This chapter explores the possibility of using the ESA scheme mentioned in the previous paragraph to automatically monitor the environment for inhaler sounds, through the use of mobile devices. The goal is to create a comprehensive system to monitor patient adherence and give practical feedback on their medication intake. This will in due course enable them to improve their quality of life and possibly reduce emergency visits.

4.1.1 Recorded Sounds

This section will explore the multiple environments in which the recorded inhaler sounds were made as well as the number of recorded inhaler sounds in the database. There were a total of eight different environments in which sounds were recorded and they can be seen in the following list:

1. Sound Chamber
2. Living Room with TV
3. Library
4. Streets (Gerrard)
5. Streets (Finch)
6. Train Station (Artificial)
7. Food Court
8. Ryerson Hallway

The aim was to obtain an ample range of different environments in which a person can use their inhalers on a regular basis. In total, there were 15 inhaler sound samples in each 10 cm increments from the microphone resulting in a sample size of 270 sound segments for each location. Each sound snippet was further truncated such that all sound files were of the same length. Overall, the total number of environment sounds and inhaler sounds recorded for the database was 2160 samples.

Note that due to legal restrictions, for the train station portion of the database, the inhaler sounds were added artificially to a different sound chamber recording which had the addition of a train station sound found on the web.

4.2 Method

A simple pipeline of the ESA scheme can be reduced to 3 main stages: the data acquisition stage, the feature extraction stage and the classification stage. Each stage was explored in-depth in previous sections.

The current workflow of the algorithm has two main stages: the training (development) and the testing stage. The main purpose of the training stage is to develop the HMM model for the classifier. Then in the testing stage, live recorded sounds are classified by the classifier based on the model obtained in the training stage.

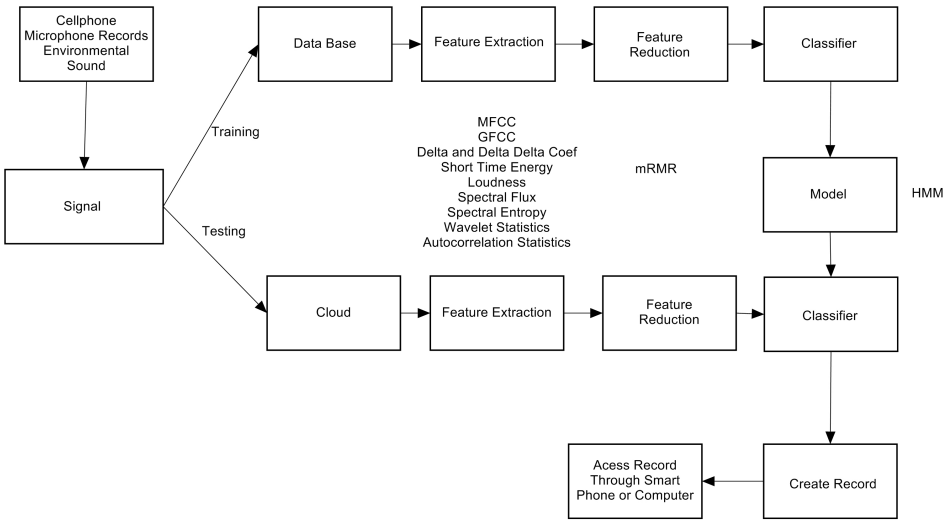

Figure 4.1: Algorithm Pipeline.

4.3 Inhaler Sound Characteristics

Before looking at inhalers in various sound environments it is crucial to identify inhaler signature characteristics. This may give us some clues as to how to structure parameters in the feature extraction step. To obtain these signature characters in a noise free environment, a sound chamber was used. Sound chambers are essentially noise canceling environments that remove both background and electromagnetic interference. These recordings are the purest inhaler sounds we can possibly obtain.

Some of the results of the sound chamber recorded inhaler sounds can be seen below for distances 10,30,50 and 90 cm. For each figure, the top most graph is the time series representation of the recording.

The figure below represents the TF map calculated using the Short Time Fourier Transform (STFT), and finally the images below are the MFCCs. Note that even at a distance of 90 cm away from the microphone we are able to identify the inhaler sound.

From Figures 4.2 - Figures 4.5, it is apparent that the inhaler sounds have a characteristic burst of energy in the 2kHz and the 4kHz frequency band. This high burst of activity also appeared on MFCC 8 and 5 which may mean that they are somewhat related.

Another important finding is that these seem to hold even after increasing the distance from the microphone. Suggesting that these features are robust with respect to the distance of signal to the microphone.

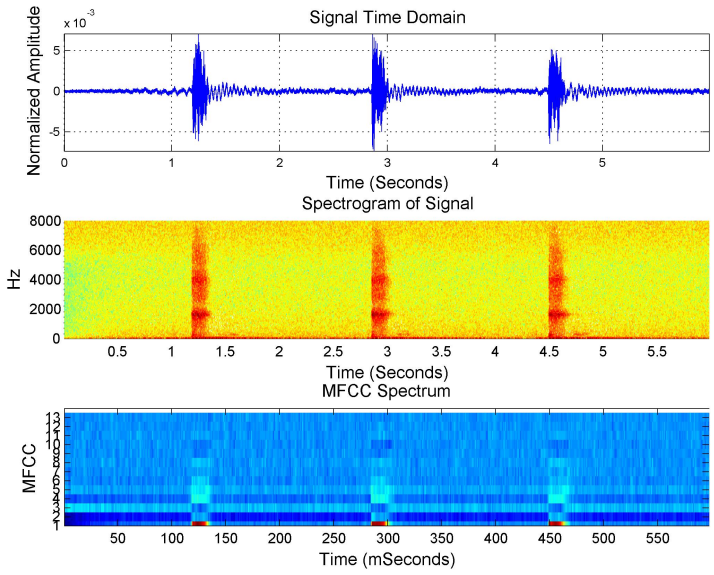

Figure 4.2: 10cm Example.

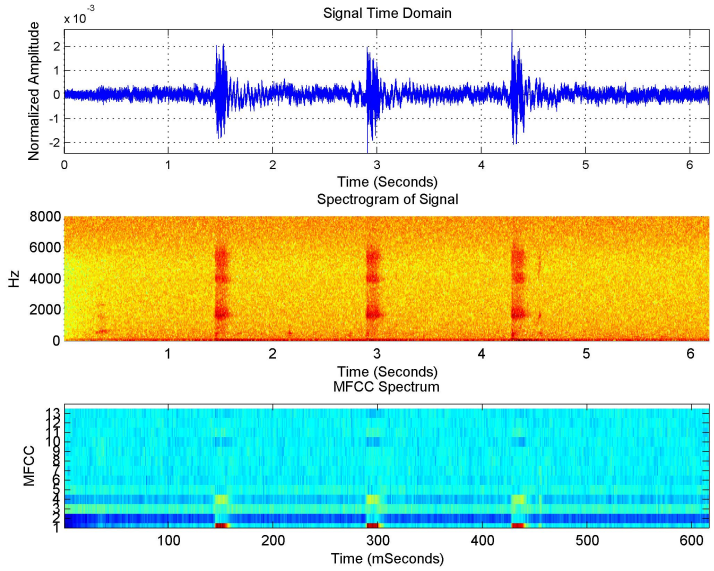

Figure 4.3: 30cm Example.

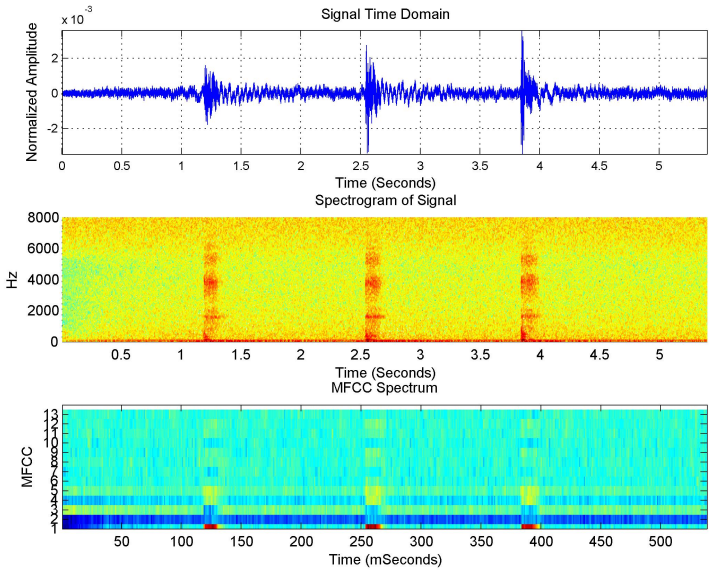

Figure 4.4: 50cm Example.

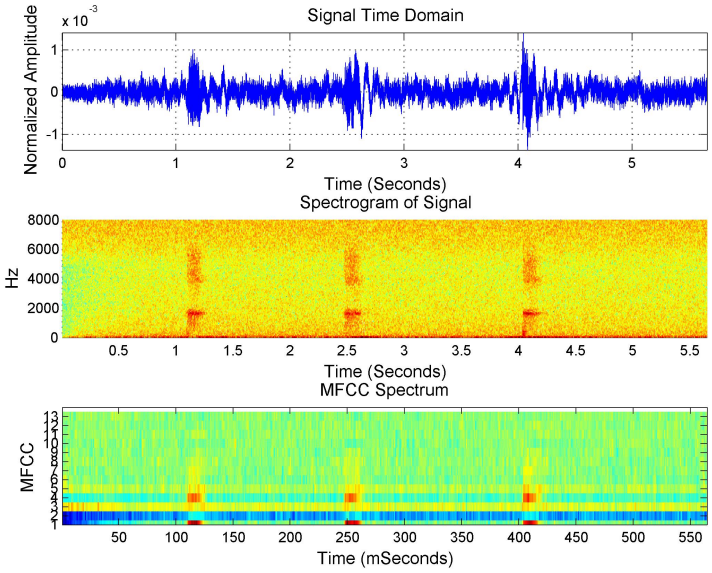

Figure 4.5: 90cm Example.

4.4 Classification

The two classifiers used were the two layer two state HMM and the two layer three state HMM. During the training stage, the Baum-Welch re-estimation algorithm was used to train the HMM. For each region, 50 percent of the inhaler sounds and 50 percent of the environment sounds were used. In addition, a log likelihood threshold was used for the testing stage and the values of the states were also used to develop likelihood criteria for each class. The remaining 50 percent of the data for each region was used for the testing stage.

The results of the classification can be seen in the next section for all the areas except for the library and the sound chamber. The reason is because in those regions the classifier obtained 100 percent classification rates. A possible explanation is that in these regions it was easy for the classifier to detect the inhalers because it was in an ideal situation where there was little to no noise present. Therefore, it is not a very good measure of the robustness of the feature, nor is it a good measure of the algorithm itself. A very interesting finding is that even in noiseless environments such as the library where typing and small talk can be heard in the background, the algorithm was still able to correctly classify inhaler sounds that were 90 cm away from the microphone. This was surprising because although the inhaler sound can be heard in the recording, it is very faint.

4.5 Results and Discussion

This work will benchmark the classification based on the classification results from the 13 MFCC features as it is currently the gold standard in ESA and speech recognition. The results for the two layer two state HMM and the two layer three state HMM can be seen in the tables below. The mRMR feature selection algorithm ranks the features in descending order from most relevant to least relevant. A problem arises when identifying the subset of these features to use as the best feature for a particular area. To solve this problem, an incremental feature HMM was used to identify the model with the lowest classification error and with the lowest number of features. Essentially 'n' number of HMM were trained, where 'n' is the number of features ranked by the mRMR algorithm. For each HMM a subset of the ranked features is used. The HMM which had the greatest classification and smallest subset of features was chosen as the model for that region. The chosen features for each region can be seen in Figure 4.1.

The table cells are colour coded to for easier reading and visualization. The yellow highlighted cells indicate the feature sets in which the location has the lowest classification error in each of the HMM designs. For example, for the food court, it was found that the first HMM design had the best classification using the mRMR reduced feature vector containing MFCC, the delta and delta delta components, and

the TF features. The red fonts indicate that the feature vector had the lowest classification error when comparing the two HMM designs. For example, for the Food Court location, the red text 23.81 in the table shows that the first HMM design had the lowest classification error compared to the second HMM design with respect to both the feature vector containing MFCC, the delta and delta delta components, and the TF features.

		HMM1 (2Layers 2States)							
	Base Classification	MFCC + delta + features		GFCC + delta + features		MFCC + GFCC + deltas		all features	
Location	MFCC	Classification Error	# of Features	Classification Error	# of Features	Classification Error	# of Features	Classification Error	# of Features
Food Court	27.78	23.81	15.00	26.19	28.00	25.40	17.00	25.40	14.00
Ryerson Hall	50.00	44.44	27.00	45.24	57.00	44.44	33.00	43.65	18.00
Living Room	0.00	0.00	9.00	0.00	12.00	0.00	10.00	0.00	9.00
Street Rush Finch	36.51	33.33	24.00	30.95	24.00	30.16	30.00	30.16	18.00
Street Traffic Gerrard	27.78	26.19	17.00	26.98	22.00	24.60	25.00	29.37	22.00
Subway	35.71	28.57	10.00	30.95	17.00	30.16	19.00	32.54	9.00

		HMM2 (2Layers 3States)							
	Base Classification	MFCC + delta + features		GFCC + delta + features		MFCC + GFCC + deltas		all features	
Location	MFCC	Classification Error	# of Features	Classification Error	# of Features	Classification Error	# of Features	Classification Error	# of Features
Food Court	38.89	26.19	21.00	26.19	28.00	26.19	32.00	26.19	34.00
Ryerson Hall	49.21	44.44	27.00	45.24	26.00	45.24	27.00	46.03	36.00
Living Room	0.00	0.00	9.00	0.00	12.00	0.00	11.00	0.00	9.00
Street Rush Finch	31.75	30.16	20.00	31.75	24.00	28.57	35.00	30.16	18.00
Street Traffic Gerrard	27.78	26.98	17.00	26.98	22.00	27.78	24.00	29.37	23.00
Subway	31.75	30.95	12.00	33.33	11.00	30.95	19.00	30.16	11.00

Table 4.1: Classification Results

Based on gross inspection of the above tables, the addition of features has improved the accuracy of classification compared to the standard MFCC features, in some cases by 7 percent. It is clear that each location has different sets of features which can discriminate the inhaler sounds and different HMM designs will generate better classification. This means that to implement this system properly, an environment classifier needs to be created to distinguish different environments from each other and then the appropriate classifier can be used to identify the inhaler sounds.

Also, regions that have high noise content seem to generate high classification errors while regions that have a relatively low noise environment have 0 classification error. This suggests that this approach is only currently suitable for low noise environments. It is also interesting to point out that the main features that discriminate the inhaler sounds from environmental sounds also seem to be the features which emulate the human ear, mainly the MFCC and GFCC and their delta derivatives to capture the

Location	Best Features
Food Court	• MFCC(10, 11, 4, 9) • MFCCΔ(1, 13, 8) • MFCCΔΔ (1, 2, 3, 4, 5, 6, 8, 11)
Ryerson Hall	• MFCC(5) • MFCCΔ(1, 2) • MFCCΔΔ (1, 2, 3, 4, 5) • GFCC(10) • GFCCΔ(1, 8) • GFCCΔΔ (1,2,3,4,5,6) • Loudness
Living Room	• MFCC(1,3,7,9,10,11,13) • MFCCΔΔ (2,3)
Street Rush (Finch)	• MFCC(1,10,11,2,5,9) • MFCCΔ(1, 2, 3, 8, 12) • MFCCΔΔ (1,10,11,13,2,3,4,5,6,) • GFCC(1,10,4,5) • GFCCΔ(1,2,3,) • GFCCΔΔ (1,2,3,4,5,6,7,9)
Street Traffic (Gerrard)	• MFCC(10,2,4,5) • MFCCΔ(1,2) • MFCCΔΔ (1,2,3,4,5,6) • GFCC(4,5) • GFCCΔ(1,2,3,6,7,8) • GFCCΔΔ (1,2,3,4,5)
Subway	• MFCC(10,9) • MFCCΔΔ (1,2,3,4) • GFCCΔ(2) • GFCCΔΔ (1,2,3,4)

Table 4.2: Reduced Features

rate of changes.

4.6 Recording Classification and Discussion

To test the algorithms in a live setting, three 18 second recordings from the living room, street and food courts were created. They range from the least to the noisiest environments. The models that were found to be the most discriminative were used in each classification. In each of the 18 second recordings nine inhaler sounds were produced with varying distances from the microphone from 10 cm to 40 cm. The results can be seen in the graphs below.

For each figure, the top most graph corresponds to the 18 sec recording where the red portion of the graph indicates when the inhaler sound was recorded. The graph in the middle is the output of the algorithm while the red portion of the graph indicates where the algorithm has detected that an inhaler sound was present. Finally, the bottom graph is a breakdown on where the inhaler algorithm has correctly or incorrectly classified the inhaler and environmental sounds. The colour scheme is as follows:

- The blue region is the correctly classified environmental sounds (True Negative)
- The green region is the correctly classified inhaler sounds (True Positive)
- The red region is the incorrectly classified inhaler sounds (False Positive)
- The magenta region is the incorrectly classified environmental sounds (False Negative)

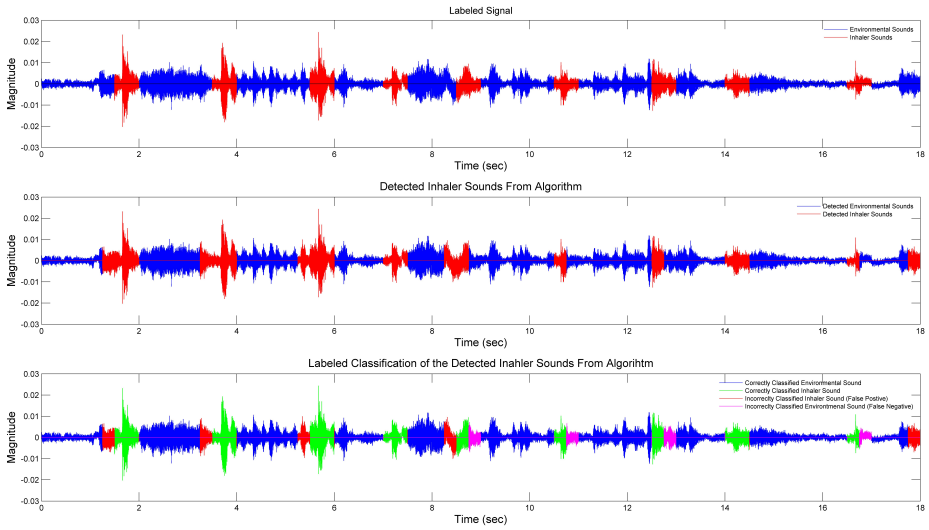

Figure 4.6: Living Room 18 Seconds Recording

In low noise environments such as a living room with the TV on (with signal to noise ratio being 4db, based on averaging the SNR of the clean inhaler sound and the environment sound), the algorithm can detect the inhaler sounds quite accurately with a very small margin of error. It has correctly identified all nine inhaler sounds and their relative location, though it also classified an environmental sound as an inhaler sound in the last 17 seconds of the recording. The cause maybe due to similar characteristics as the inhaler sounds.

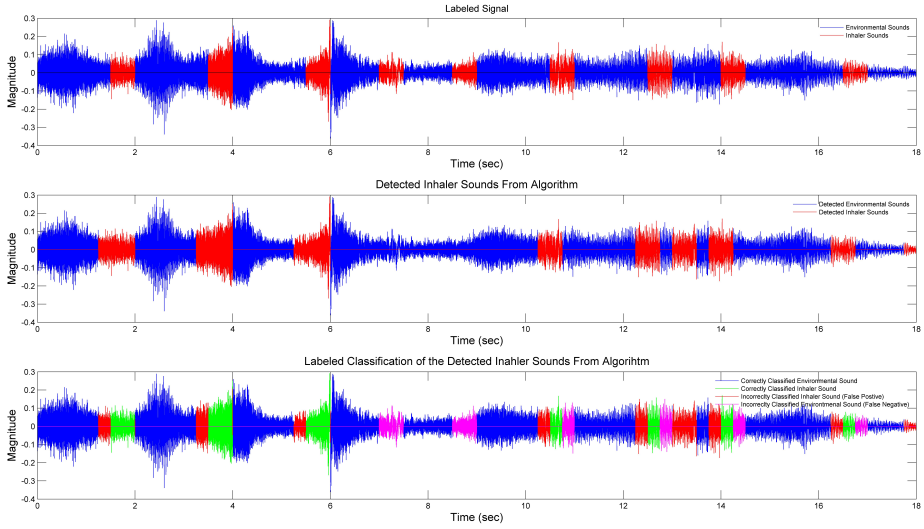

Figure 4.7: Street 18 Seconds Recording

In medium noise environments such as a side street in which cars occasionally pass by (with SNR being 21 db), the algorithm can still detect the inhaler sounds but not as accurately. In the above example, out of the nine inhaler sounds, two were misclassified and two were false positives (i.e. they were classified as inhaler sounds even though they were environmental sounds).

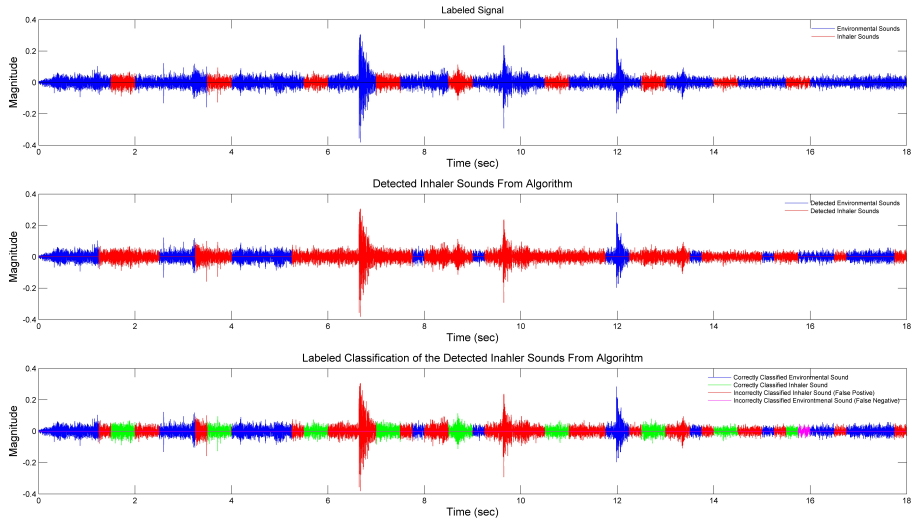

Figure 4.8: Food Court 18 Seconds Recording

In high noise environments such as the food court (with SNR being 25 db) it was very difficult for the algorithm to correctly detect and locate the inhaler sounds. Many of the environmental sounds were classified as inhaler sounds as seen in the second graph. This may be due to the fact that many of the sounds were not used as training for the HMM classifier, because this recording was done at a later date.

The results laid out in this section show that to improve classification, different sounds that occur in the environment must be trained in the classification algorithm. In this way, the classification algorithm will not confuse similar sounding sounds in the environment with the inhaler made sounds.

Because this algorithm needs to run on the phone as an application, it is important to do some power and computation analysis. An easy method is to see how long it takes to calculate a half second recording running on a normal PC and then compare the specs of the PC to that of a cellphone. On average it takes about 0.14 secs for a half a second recording to be classified as either an environmental or an inhaler signal. This number was generated by running the algorithm on MATLAB 2011b on a computer with an 8mb of ram and running a 3.3GHz dual core Intel i3-2120 CPU. Comparing these specs to the HTC One (M8) running a Qualcomm Snapdragon 801 quad-core CPU. It is easy to see that performance will only increase to near real time monitoring of the environment with multi-threading capabilities on smart phones. In terms of power consumption, it is not practical for the app to be recording and processing sounds indefinitely. Instead, a system in which the gyroscope may be utilized to only record if the person is stationary. In this way, the app will only activate in certain conditions and consume less power.

Chapter 5

Conclusions and Future Work

This work presented a novel AD application for respiratory sound analysis for easy detection of crackle sounds by the clinicians. Though the sounds have been made audible by audification, the sonification technique is by far the most appropriate AD technique to use for respiratory sound analysis. Though parametric mapping of the EMD, wavelet and bump source, and TF map of MP sonification enhanced the ability for the user to identify crackles, because sound is very subjective and is perceived differently among users this type of analysis cannot be incorporated smoothly into the clinicians workflow. Instead, further training of this technique is needed for current physicians but for aspiring doctors this tool is ideal for future diagnosis of early respiratory diseases.

For the AD analysis of respiratory sounds all the current implemented and developed algorithms do not take into account where the auscultation sounds originate. Depending on the auscultation site, the crackle may be outside the audible domain, or so faint that the extraction algorithms explored cannot separate the two. Therefore, a study must be done to see the effects different auscultation sites have on the separation level as well as the sonification or audification level. These methods have to be tested on actual physicians to see how accurate they are and how helpful they are in diagnosing patients with abnormalities. Another approach to sonification is using the envelope of the IMF to be the envelope of the sounds being generated by the MIDI instead of a direct mapping of the TF spectrum.

Secondly, this work presented a novel application of ESA to inhaler made sound detection for asthma medication adherence. The features used in the ESA techniques were used mainly to emulate the human auditory system as well as characterise the noise like property of the inhaler sound. It was found that in low to medium noise environments, this algorithm can detect inhaler sounds with a small margin of error while in high noise environments the algorithm had a hard time differentiating inhaler sounds from

environmental sounds. To improve classification, we can train the HMM classifier with various sounds heard in the environment so it learns that these are part of the environment and are not an inhaler sounds. Also by denoising the signal as a pre-processing step classification can be further improved.

Future studies can also incorporate the gyroscope of the phone to remove false positives. If the gyroscope shows that the person is walking, then the algorithm should not run because the person would not be using their inhalers in the first place. This method of only running the algorithm when the person is stationary can also reduce power consumption of the smart phone.

List of Acronyms

AD - Audio Display

ADSR - Attack Decay Sustain Persuit

ASC - Asthma Society of Canada

COPD - Chronic obstructive pulmonary disease

DCT - Discrete Cosine Transform

DFT - Discrete Fourier Transform

EMD - Empirical Mode Decomposition

ESA - Environmental Sound Analysis

FFT - Fast Fourier Transform

FIR - Finite Impulse Response

GFCC - Gamma Frequency Cepstral Coefficients

GUI - Graphical User Interface

HCI - Human to Computer Interface

HHS - Hilbert Huang Spectrum

HMM - Hidden Markov Model

HS - Hilbert Spectrum

ICA - Independent Component Analysis

IMF - Intrinsic Mode Functions

IWPT - Inverse Wavelet Packet Transforms

MFCC - Mel Frequency Cepstral Coefficient

MP - Matching Pursuit

mRMR - Maximum Relevance and Minimum Redundancy

PCA - Principal Component Analysis

PHAC - Public Health Agency of Canada

PV - Phase Vocoder

SNR - Signal to Noise Ratio

STFT - Short Time Fourier Transform

TF - Time-Frequency

WBS - Wavelet and Bump Source

WPT - Wavelet Packet Transformation

References

[1] (2014, Sept) Stethoscope-2.png. [Online]. Available: http://upload.wikimedia.org/wikipedia/commons/d/d2/Stethoscope-2.png

[2] (2014, Sept) digitalstethoscope.jpg. [Online]. Available: http://www.rechargelifecorp.com/img/digitalStethoscope.jpg

[3] (2014, Sept) Human respiratory system. [Online]. Available: http://suumsleepclinic.tistory.com/450

[4] (2014, Sept). [Online]. Available: http://www.asthma.ca/images/global/lungsNormal.jpg

[5] (2014, Sept). [Online]. Available: http://www.asthma.ca/images/global/lungsAsthmatic.jpg

[6] (2014, Sept). [Online]. Available: http://www.cutandjacked.com/The-Ins-And-Outs-Of-Asthma

[7] (2014, Sept). [Online]. Available: http://upload.wikimedia.org/wikipedia/commons/thumb/d/d2/Anatomy_of_the_Human_Ear.svg/800px-Anatomy_of_the_Human_Ear.svg.png

[8] J.-L. Beaumont, *Lexamen clinique respiratoire avec cassette audio des bruits respiratoires normaux et anormaux*, 2nd ed. Gestion J.L. Beaumont, 1999.

[9] (2014, Sept). [Online]. Available: http://msis.jsc.nasa.gov/sections/section03.htm

[10] (2014, Sept) Public health agency of canada. [Online]. Available: http://www.phac-aspc.gc.ca/publicat/2007/lbrdc-vsmrc/index-eng.php#tphp

[11] Z. Moussavi, "Respiratory sound analysis [introduction for the special issue]," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 26, no. 1, pp. 15–15, Jan 2007.

[12] "Long-acting -agonists in asthma: an overview of cochrane systematic reviews," *Respiratory Medicine*, vol. 99, no. 4, pp. 384 – 395, 2005.

[13] B. Bender, H. Milgrom, and C. Rand, "Nonadherence in asthmatic patients: Is there a solution to the problem ?" *Annals of Allergy, Asthma & Immunology*, vol. 79, no. 3, pp. 177 – 187, 1997.

[14] (2014, Sept) Chronic respiratory diseases. [Online]. Available: http://www.who.int/respiratory/asthma/en/

[15] L. Heaney and R. Horne, "Non-adherence in difficult asthma: time to take it seriously." *Thorax*, vol. epub, no. 3, pp. 268–270, 2011.

[16] D. Price, S. Bosnic-Anticevich, A. Briggs, H. Chrystyn, C. Rand, G. Scheuch, and J. Bousquet, "Inhaler competence in asthma: Common errors, barriers to use and recommended solutions," *Respiratory Medicine*, vol. 107, no. 1, pp. 37 – 46, 2013.

[17] M. Holmes, S. D'arcy, R. Costello, and R. Reilly, "Acoustic analysis of inhaler sounds from community-dwelling asthmatic patients for automatic assessment of adherence," *Translational Engineering in Health and Medicine, IEEE Journal of*, vol. 2, pp. 1–10, 2014.

[18] (2014, Sept) Asthma. [Online]. Available: http://www.lung.ca/diseases-maladies/asthma-asthme/treatment-traitement/medications-medicaments_e.php

[19] H. Pasterkamp. (2014, Sept) The r.a.l.e. repository. http://www.rale.ca/Default.htm.

[20] J. Chan, H. Ma, T. Saha, and C. Ekanayake, "Self-adaptive partial discharge signal de-noising based on ensemble empirical mode decomposition and automatic morphological thresholding," *Dielectrics and Electrical Insulation, IEEE Transactions on*, vol. 21, no. 1, pp. 294–303, February 2014.

[21] C. Qu and Y. Wang, "Natural frequencies identification of subgrade for high-speed railway based on empirical mode decomposition," pp. 4181–4184, June 2011.

[22] S. Shukla, S. Mishra, and B. Singh, "Empirical-mode decomposition with hilbert transform for power-quality assessment," *Power Delivery, IEEE Transactions on*, vol. 24, no. 4, pp. 2159–2165, Oct 2009.

[23] Z. Liu, H. Wang, and S. Peng, "Texture classification through directional empirical mode decomposition," vol. 4, pp. 803–806 Vol.4, Aug 2004.

[24] A. Karagiannis, L. Loizou, and P. Constantinou, "Experimental respiratory signal analysis based on empirical mode decomposition," p. 15, Oct 2008.

[25] T. Sakai, M. Kato, S. Miyahara, and S. Kiyasu, "Robust detection of adventitious lung sounds in electronic auscultation signals," pp. 1993–1996, Nov 2012.

[26] T. Sakai, H. Satomoto, S. Kiyasu, and S. Miyahara, "Sparse representation-based extraction of pulmonary sound components from low-quality auscultation signals," pp. 509–512, March 2012.

[27] T. Rutkowski, A. Cichocki, and D. Mandic, "Information fusion for perceptual feedback: A brain activity sonification approach," in *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, D. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, Eds. Springer US, 2008, p. 261273. [Online]. Available: http://dx.doi.org/10.1007/9780387743677_14

[28] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[29] C. Clavel, T. Ehrette, and G. Richard, "Events detection for an audio-based surveillance system," pp. 1306–1309, July 2005.

[30] T. Brandes, "Feature vector selection and use with hidden markov models to identify frequency-modulated bioacoustic signals amidst noise," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 6, pp. 1173–1180, Aug 2008.

[31] D. Mathias, A. Thode, S. Blackwell, and C. Greene, "Computer-aided classification of bowhead whale call categories for mitigation monitoring," pp. 1–6, Oct 2008.

[32] H. Itoh, T. Takiguchi, and Y. Ariki, "Event detection and recognition using hmm with whistle sounds," pp. 14–21, Dec 2013.

[33] D. Hollosi, J. Schroder, S. Goetze, and J.-E. Appell, "Voice activity detection driven acoustic event classification for monitoring in smart homes," pp. 1–5, Nov 2010.

[34] A. Dufaux, L. Besacier, M. Ansorge, and F. Pellandini, "Automatic sound detection and recognition for noisy environment," 2000.

[35] O. Uribe, H. Meana, and M. Miyatake, "Environmental sounds recognition system using the speech recognition system techniques," pp. 13–16, Sept 2005.

[36] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 4, pp. 357–366, Aug 1980.

[37] X. Valero and F. Alias, "Gammatone wavelet features for sound classification in surveillance applications," pp. 1658–1662, Aug 2012.

[38] F. Su, L. Yang, T. Lu, and G. Wang, "Environmental sound classification for scene recognition using local discriminant bases and hmm," pp. 1389–1392, 2011.

[39] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.

[40] A. TorresJimenez, S. CharlestonVillalobos, R. GonzalezCamarena, G. ChiLem, and T. Aljama-Corrales, "Asymmetry in lung sound intensities detected by respiratory acoustic thoracic imaging (rathi) and clinical pulmonary auscultation," p. 47974800, Aug 2008.

[41] (2014, Sept) Respiratorysounds. [Online]. Available: http://adesa4respsig.wordpress.com/