Escuela Politécnica Superior de Linares

Masters In Telecommunication Engineering

Universidad de Jaén

Escuela Politécnica Superior de Linares

WHEEZING SOUND SOURCE SEPARATION APPLIED TO SINGLE CHANNEL RESPIRATORY AUDIO MIXTURES

Alumno: Sindhusha Jeeru

Tutor : Prof. Fco Jesus Cañadas Quesada
Tutor : Juan de la Torre Cruz
Depto.: Department of Telecommunication Engineering

September, 2020

TABLE OF CONTENTS

FIGURES INDEX

TABLES INDEX

1 ABSTRACT

Listening to lung sounds through stethoscopes has been an ancient technique followed by physicians since 1816 for general examinations of patients who are critically ill with respiratory disorders. However lung auscultation has been an important process followed by clinicians to take decisions on patient's respiratory health by listening to the wheezing sounds produced during human breathing. Wheezing sounds are adventitious sounds that occur if there is any obstruction of airflow in the human breathing. The wheezing sounds leave very important information about respiratory disorders related to Chronic Respiratory diseases like Asthma and Chronic Obstructive Pulmonary Disease (COPD). Although these diseases are not curable, they can be prevented from worsening the condition by close monitoring of the symptoms. Wheezing is considered as one of the main symptoms that have to be regularly monitored to know the criticality of the lung disorder. Therefore early identification of occurrence of wheezing sounds during breathing could prevent serious exacerbations. However, auscultation is an improper treatment and referrals accumulate an increased time and monetary cost. Moreover training physicians is a challenging task because of varying perception of sound by different people. For examination of some respiratory diseases there is also need for highly professional pulmonary experts which has been a challenging task. As a consequence of these challenges, there is a demand for tools that automatically separates wheezing sound from respiratory sounds that provides physicians with good audio quality wheezing sounds.

Therefore our master thesis study is to provide a novel wheezing source separation algorithm that exclusively removes interfered respiratory sounds from monaural or single channelled respiratory audio mixtures and leaves the separated wheezing sounds. Our proposed study is an improved version of the Non-Negative Matrix Partial Co-Factorization (NMPCF) that resolves the drawbacks that occur in other conventional NMPCF approaches by tending to capture both the spectral and temporal repetitive components of the interfered respiratory sounds from respiratory audio mixture and preserves the separated wheezing sound for future diagnosis. The other conventional approaches like Spectral NMPCF (S-NMPCF) captures only spectral components of the

respiratory signal whereas Temporal NMPCF (T-NMPCF) captures only the repetitive events of the target signal from the respiratory audio mixture. That is why our improved version is named as Spectral and Temporal NMPCF (ST-NMPCF) for wheezing source separation. In our proposed approach we consider spectrogram matrices of respiratory training signal(which contains solo-respiratory sounds) and spectrogram matrices of partitioned column blocks of a respiratory mixture signal (a composition of respiratory and wheezing sounds) and apply ST-NMPCF approach on these target matrices. By doing this, the partitioned column blocks of the respiratory audio mixture and respiratory training signal get jointly decomposed and share a factor matrix partially, in order to determine common basis vectors that capture the spectral and temporal characteristics of respiratory sounds. The Common basis vectors learned by ST-NMPCF capture spectral patterns of respiratory sounds since they are shared in the decomposition of the respiratory training data matrix and temporal patterns of respiratory sounds are accommodated because repetitive characteristics are captured by factorizing column-blocks of mixture spectrograms. Thus resulting in separation of spectral and temporal respiratory sounds from the respiratory audio mixture and leaving behind the wheezing sounds.

2 INTRODUCTION

2.1 Motivation

Respiratory diseases are treated as a world health threat because of the increasing death rates caused due to these diseases. Chronic Respiratory Diseases (CRDs), such as Asthma and Chronic Obstructive Pulmonary Disease (COPD), are mentioned as leading causes of deaths worldwide. According to World Health Organization reports it is estimated that more than 235 million people suffer from asthma worldwide, causing literally 300,000 deaths per year [1]. Asthma is regarded as such a lung disorder that affects all ages, races and ethnicities, though there exists wide variation in different countries and in different groups within the same country. It is the most common chronic disease that affects 14% of all children globally and is more severe in children living in non-affluent countries [2]. Moreover COPD is regarded as another respiratory threat that causes even more deaths compared to asthma. It is estimated that more than 250 million cases are reported annually which are further resulting in 3 million deaths globally [3]. It is stated that COPD is going to become the one of the third leading causes of deaths worldwide by 2030. The first two being the ischemic heart disease and cerebrovascular disease [4].

Besides the death rate, if we talk about the economic impact caused due to respiratory diseases, these are highly influencing the economic situation of the countries. As an illustration, in the UK they cost the National Health Service approximately 3 and 2 million GBP, for asthma and COPD, respectively [5].

Asthma and COPD are diseases which are not curable but upon continuous monitoring of symptoms they are supposed to reduce the acquainted risks and slow down the disease worsening. These diseases are generally characterized by symptoms like wheezing, coughing, and breathlessness. These symptoms affect the individual on a daily basis and they are supposed to worsen while the individual is involved in physical activities. They are often worse at night time and early mornings [1] causing sleep disruptions [5] and they are supposed to consequently reduce the quality of life of those affected. The worsening

of symptoms in Chronic Respiratory Diseases (CRDs) is known as exacerbations. Exacerbations can be life-threatening and can permanently alter respiratory system structures. In some countries, such as the UK, they are also one of the leading causes of emergency hospital admissions [6]. Henceforth it is very important to detect these symptoms at an early stage to improve the quality of patient's health as well as to low the death rate.

The reference papers [8] [9] explains about the importance of identification and monitoring of these key symptoms at their early stage is critically necessary for management of CRDs. Among these symptoms wheezing is considered as one of the key symptoms that have to be monitored regularly to know the actual situation of the patient's lung health. The main reason behind monitoring the wheezing sound is that they are continuous abnormal breathing sounds with a specific tone [10], and they are usually considered as a clinical sign in pulmonary medicine that reflect the degree of airway obstruction caused due to narrowing of airways [11]. The occurrence of wheezing sound indicates that the patient is suffering to get sufficient amounts of air to maintain normal breathing, resulting in dyspnoea, asphyxia or other life-threatening situations [12]. Therefore, it is important to detect and monitor wheezing sounds to provide adequate medical treatment for patients at their early stages to improve their lung health and increase their span of life.

These sounds vary from person to person. So listening and examining them is a challenging task for physicians. Moreover clinicians follow the old auscultation process where they use stethoscope as a main tool to listen to the wheezing sounds. The whole process is performed in a quiet environment and ideally with the patient in a still position where the clinician auscultates systematically the anterior, posterior, and lateral lung fields by placing the stethoscope over the patient chest. He then compares sounds heard on one side to sounds heard in the same location on the opposite side and check for bilateral asymmetry of sound amplitude and that asymmetry indicates disease. This process restricts the duration and flexibility of monitoring. Though this process is simple and convenient, it is completely dependent on the experience of the physicians, variability of the human auditory system and specifications of the stethoscope [13] [14]. It is also a fact that the physician's audibility to perceive the lung sounds reduces if he continuously

monitors throughout the day for long hours through the stethoscope. This affects the ability to effectively diagnose the individual's lung health. Moreover Wheezing sounds are heard along with respiratory sounds because they are superimposed on the respiratory sounds. This interference of respiratory sounds don't allow the physicians to have a clear cut audibility of wheezing sounds which results in loss of information of the wheezing sounds. This tends to lead to poor treatment of lung disorders. Therefore it is necessary to have researches or studies on source separation algorithms which separate the respiratory sounds from the signal composed of both respiratory and wheezing sounds and leaving the isolated wheeze sounds. This helps the physicians in getting the complete information regarding lung disorders through the isolated wheezing sounds.

2.2 Structure of TFM

This document is organized as follows:

The further part of the document in Section 2 explains the Auscultation process which was the initial step followed to get the lung sounds, a brief note on terminology of breath sounds and we ended up the section with state-of-art-methods explaining the novelty of this thesis.

Section 3 describes the main objectives of our thesis study. This section mainly focuses on the actions that have to be undertaken in order to achieve the results of the proposed study.

Section 4 is about Materials and Methods where we explain to the readers in detail about the proposed approach and the guides we followed to reach the proposed approach.

Section 5 details about the obtained experimental results with the explanation of databases creation and its features. In this section we also report about results obtained from testing and optimization process along with comparison results.

In Section 6 and Section 7 we are going to discuss and conclude some important observations that are drawn from our thesis study

Section 8 and Section 9 informs about the references and appendix

2.3 Auscultation

Auscultation is a medical term used by clinicians which defines about the action of listening to sounds from the heart, lungs, or other organs, typically with a stethoscope, as a part of medical diagnosis. In the same way pulmonary auscultation or auscultation of breath sounds represents listening to sound produced by lungs. Auscultation is regarded as the one of the oldest, safe-to-perform, and inexpensive techniques used by the physicians to diagnose various pulmonary related diseases. With the invention of traditional stethoscope by René Laennec in 1816 to till date stethoscope has been widely used tool by clinicians to examine the patient's health. The advancement in electronics, materials and construction gave rise to development of commercial stethoscopes. Presently, two different types of stethoscopes are available on the market: acoustic and electronic stethoscopes. Choosing the optimal stethoscope is vital for accurate auscultation and patient care because poor auscultation can adversely affect patient outcomes and increase healthcare expenditures. [83] [84] Acoustic stethoscopes are non-electronic tools and the main advantages of acoustic stethoscopes are robustness and comfort in working environment. But these are not ideal to use because the sound level is very low because sound energy travels internally from the patient, through a conductor (the stethoscope), and to the listener (the clinician). This type of transmission results in poor quality of sound and results in loss of hearing aid of the low frequency sounds. According to the comments of the physician and nurses the specific limitations of acoustic stethoscopes are: (1) the lack of amplification of the sounds, (2) more attenuation for the higher frequency sounds, and (3) the high pressure applied on the ears by some models of stethoscope to isolate the lung sounds from ambient sounds. But the electronic stethoscopes utilize advanced technology to overcome these low sound levels by electronically amplifying the lung sounds. [83] [84] Electronic stethoscopes converts the acoustic sound waves obtained through the chest piece into electronic signals which then transmits through uniquely designed circuitry installed in it and process it for optimal listening. They also improve accuracy, efficiency of pulmonary auscultation and also allow sophisticated computerised analysis. Some electronic stethoscopes have the advantages of digitizing the audio signal. Although the electronic stethoscope partly solved the

limitations of the acoustic stethoscopes, it also introduced other limitations like (1) interference of electronic noise (2) heavier and less mobile (3) sensitivity to impact, manipulation and ambient noises.

In general auscultation of breath sounds include evaluation of various parameters. These evaluation parameters mostly include quality of sounds, presence of additional sounds i.e. other than respiratory sounds, their timing within the respiratory cycle, their amplitude, and intensity of the sounds [72]. The information about the quality of the sounds is described in terms of vesicular, bronchovesicular, and bronchial. Whereas the second parameter regarding the interference or presence of additional sounds implies the physicians write a note of occurrence of adventitious sounds like wheezing sounds, crackles sounds, and pleural rib by listening to the sounds through a stethoscope [73], [74]. The timing within the respiratory cycle refers to respiratory events i.e. they refer to inspiration and expiration phases of the breath sounds. In our document whenever we refer to a respiratory event it implies the inspiration and expiration phase of the breathing on a whole. The final parameter that the physician has a note on is the amplitude of the sounds, the sounds include both normal and additionally interfered sounds. This is subjectively evaluated by clinician and based on this information they usually describe the sounds in three categories: normal, decreased (or reduced or diminished) and abnormal [81]. The clinician confirms the breathing sounds as normal if the sounds heard from anterior and posterior surface of the chest do not have any asymmetry and they are not interfered with adventitious sounds or abnormal sounds. Likely, the physician confirms the breathing sounds as decreased if the breathing sounds are inaudible (absent) or reduced in volume compared with other areas of the lungs when examined with a stethoscope. Absent breath sounds are often caused by major or minor airway obstruction that results in no air flow [82]. Similarly the breathing sounds are categorised as abnormal if the clinicians hear adventitious sounds or abnormal sounds while examining the breathing of the patient. There are a number of different terms used to describe abnormal or adventitious breath sounds, and these are very confusing. Some are heard with a stethoscope (auscultation), but some may be heard without. These sounds differ based on whether they are predominant during inspiration or expiration, in the quality of the sounds, and more.

Normally auscultation is carried out in a quiet room, where the patient will be asked to take deep breaths through the open mouth. Using the diaphragm of the stethoscope the physician listens for the quality of the breath sounds, the intensity of the breath sounds, and the presence of adventitious sounds. The physician carries out this process of listening to the breath sounds by placing the stethoscope on various sites of patient's chest surface on front and back parts. The physician listens to at least one complete respiratory event at each site he performs the test. But the application of the traditional stethoscope in research studies has been limited due to the variability of human audition to perceive the sounds, subjectivity and expertise in the interpretation of lung sounds and also difference in labelling of the sounds. There was a study made to measure the accuracy of lung auscultation practiced by physicians and medical students [77], and they concluded and confirmed that there is an ambiguity in identification and interpretation of sounds from person to person.

Compared to manual subjective auscultation, the automated or computerized analysis of lung sounds [75] (which is otherwise called as electronic auscultation) gained wide popularity because its objective nature of analyzing breath sounds has greatly improved the diagnostic process. Unlike other diagnostic methods the objective evaluation of breath sounds is based on reproducing the data that can be stored and analyzed in terms of quantitative parameters which the human ears does not meet through manual auscultation process. To increase the usefulness of clinical diagnosis and attempts to learn the physics and physiology behind the breath sounds the time domain and frequency domain methods were introduced. In spectral representation, abnormal lung sounds, such as wheezes and crackles, show recognizable patterns, (***Figure 2.3.1 and Figure 2.3.2***) that are useful in the identification of

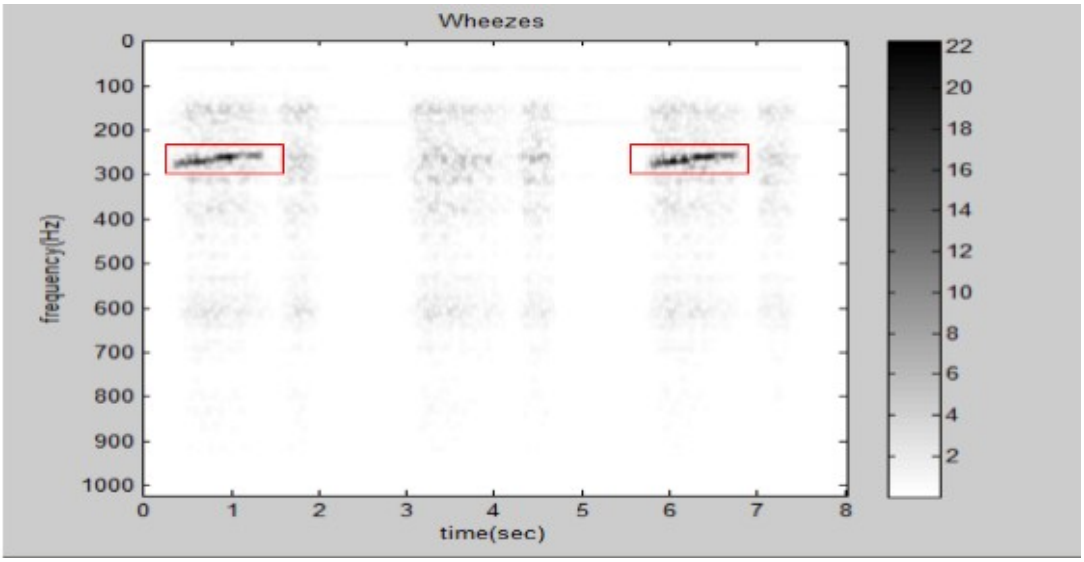


Figure 2.3.1 Spectrogram of respiratory signal representing two wheezes in red colour box

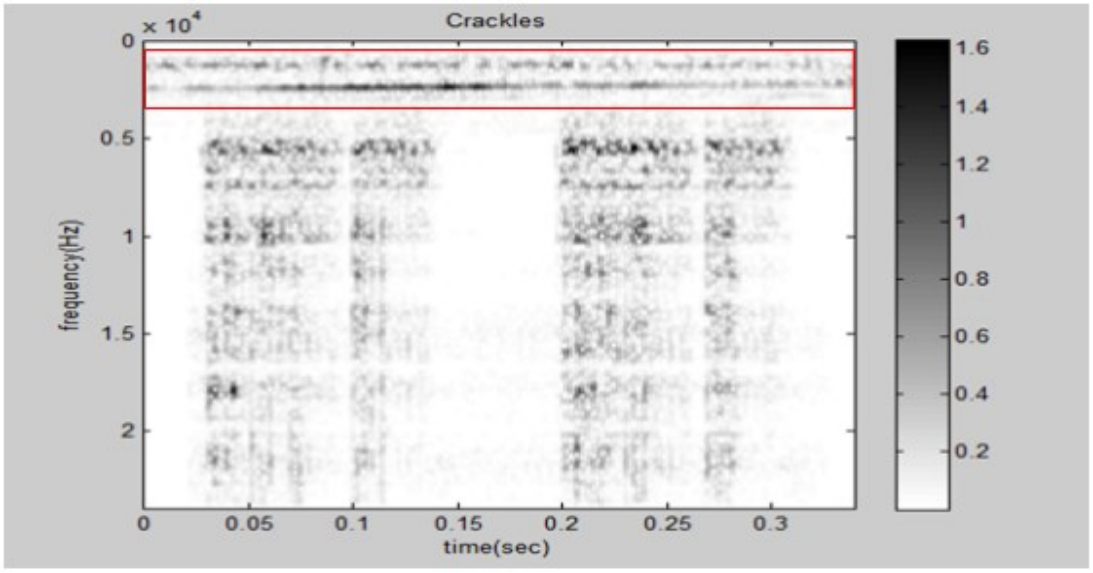


Figure 2.3.2 Spectrogram of respiratory signal with crackles represented in red colour box

these sounds. In the reference paper [76] the authors made a statistical study on explaining about the significance of spectral signatures of the lung sounds. They also reported in their document on how several researches made on lung sound analysis used spectral information as preliminary investigation to proceed with the works.

2.4 Respiratory sounds terminology

Respiratory sounds refer to the specific sounds generated when there is movement of air through the respiratory system. The lungs produce three categories of sounds, which clinicians listen during auscultation: breath sounds, adventitious sounds, and vocal resonance. Breath sounds are normal lung sounds heard through the chest wall with the use of a stethoscope, rather than audible breathing through the mouth. Breath sound has three characters; frequency, intensity, and timbre or quality; which helps in differentiating two similar sounds. Frequency measures the number of the sound waves or vibrations per second and is measured objectively. It is measured in hertz (Hz). Pitch is the subjective perception of a sound's frequency. Pitch depends on the frequency and is within 5 Hz of the frequency usually [79]. The human ear can perceive sound waves over a wide range of frequencies, ranging from 20 to 20,000 Hz [80]. Amplitude is related to the energy of sound waves and is measured by the height of sound waves from the mean position. Loudness is the subjective perception of amplitude. Quality or timbre is an important property of sound that differentiates two sounds with the same pitch and loudness. Description and classification of the sounds usually involve auscultation of the respiratory event of the breathing sound, noting both the pitch (typically described as low, medium or high) and intensity (soft, medium, loud or very loud) of the sounds. Based on all these properties lung sounds are divided into two categories:

1. Normal breath sounds
2. Abnormal or adventitious breath sounds

The reference paper [78] published in the European Respiratory Journal, provides more accurate definitions and information of currently used terms in pulmonary auscultation domain and sound analysis; We briefed few of them by using this paper as a guide.

2.4.1 Normal Breath Sounds

Normal Breath Sounds are categorized into bronchial, vesicular, or bronchovesicular. Based on the anatomical position of the Auscultation process done, these have different acoustic properties. The normal lung

sound spectrum is devoid of discrete peaks and is not musical; this can be observed in ***Figure 2.4.1, Figure 2.4.2, and Figure 2.4.3.***

1. **Bronchial Sounds:**

Bronchial sounds are present over the large airways near the second and third intercostals spaces. These sounds are more tubular and hollow sounding than vesicular sounds. They are typically louder and higher in pitch than vesicular breath sounds. That is why these sounds are also termed as tubular sounds. They normally arise from the tracheobronchial tree. Loud, harsh, and high pitched bronchial sounds are typically heard over the trachea or at the right apex. They are predominantly heard during expiration. If heard in other areas of the lung, bronchial sounds are abnormal. For bronchial sounds the expiratory phase is greater than inspiratory phase. The expiratory pitch is high and intensity is loud. Hollow, tubular sounds that are lower pitched. ***Figure 2.4.1*** represents the spectrogram representation of the bronchial sound that is recorded from the breathing of the healthy individual via stethoscope

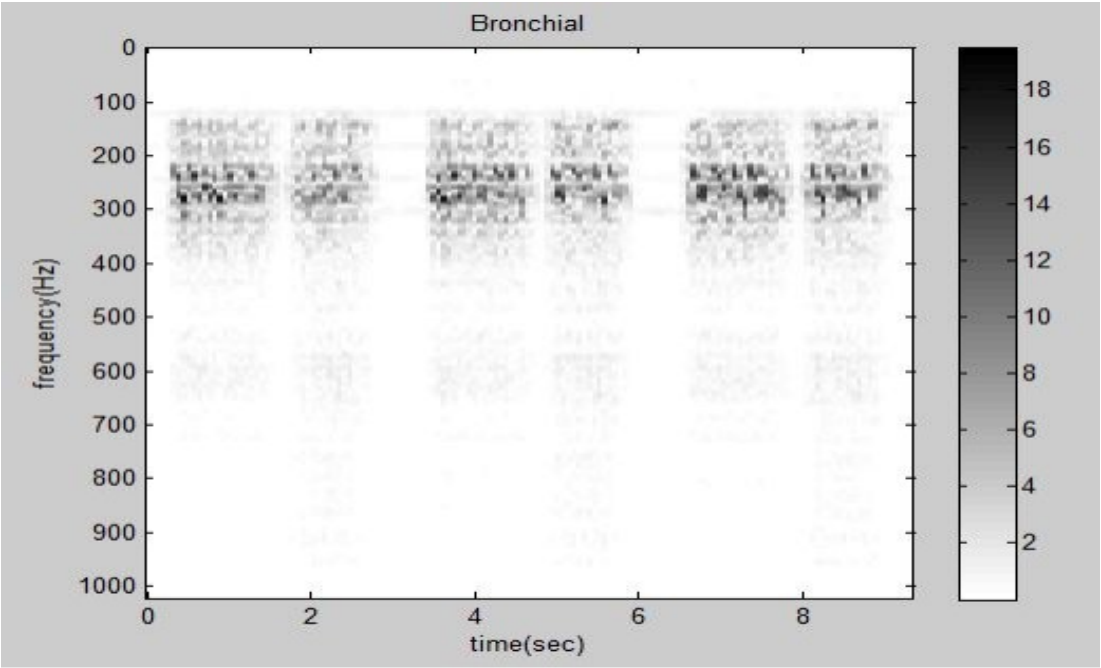


Figure 2.4.1 Spectrogram representation of bronchial sound recorded from a healthy person

2. **Bronchovesicular Sounds**

Bronchovesicular sounds can be heard during inspiration and expiration and have a mid-range pitch and intensity. They are commonly heard over the upper third of the anterior chest. Increased intensity of bronchovesicular sounds is most often associated with increased ventilation or pulmonary consolidation. The inspiratory and expiratory phases in bronchovesicular sounds are roughly equal in length. They reflect a mixture of the pitch of the bronchial breath sounds heard near the trachea and the alveoli with the vesicular sound. ***Figure 2.4.2*** represents the spectrogram representation of the bronchovesicular sound that is recorded from the breathing of the healthy individual via stethoscope

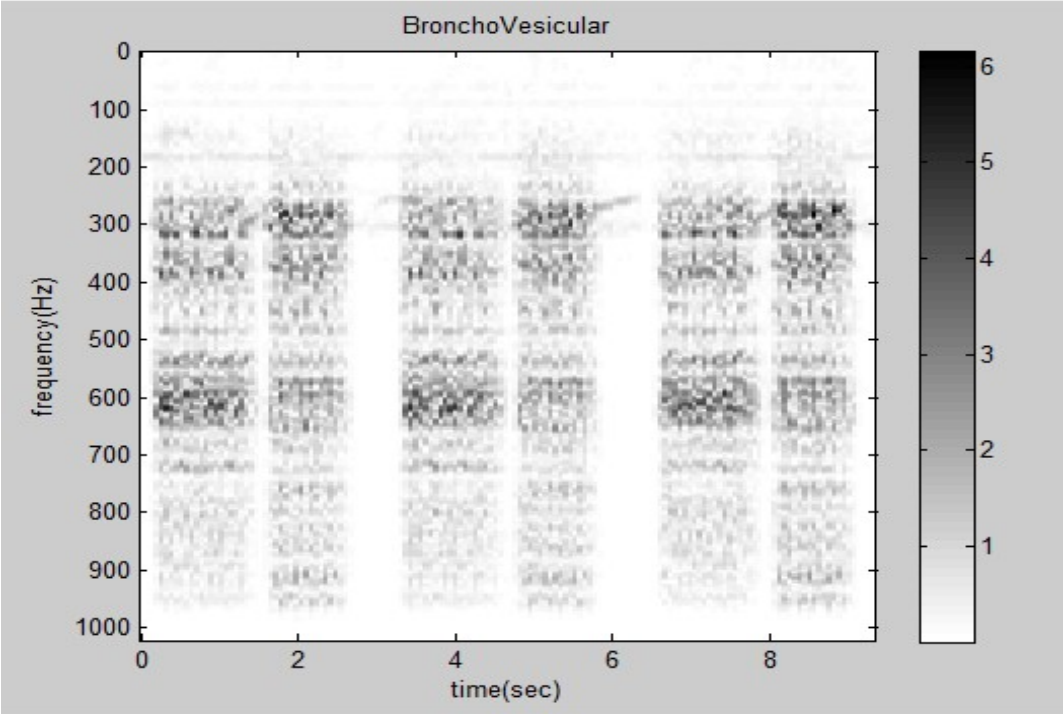


Figure 2.4.2 Spectrogram representation of Bronchovesicular sound recorded from a healthy person

3. Vesicular

Vesicular breath sounds are caused by the normal movement of air through the bronchioles during inspiration and expiration. They are heard over the periphery of the lung field. These sounds are the result of attenuation of breath sounds produced in the bronchi at the higher region of the lungs. They are highly variable in intensity depending on the species, ventilation, and body condition. These sounds are appreciated

especially well at the posterior lung bases. ***Figure 2.4.3*** represents the spectrogram representation of the vesicular sound that is recorded from the breathing of the healthy individual via stethoscope

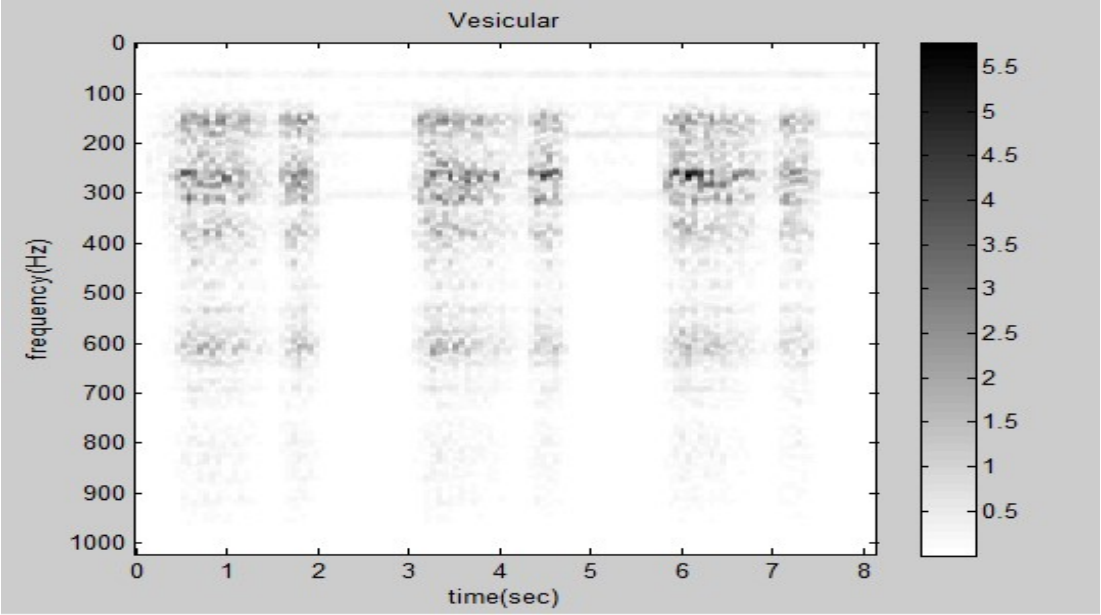


Figure 2.4.3 Spectrogram of vesicular sounds recorded from a healthy person

2.4.2 Abnormal or Adventitious Breath Sounds

Abnormal breath sounds are those that include the absence or reduced intensity of sounds where they should be heard or, by contrast, the presence of sounds where there should be none, as well as the presence of adventitious sounds. Adventitious sounds refer to sounds that are heard in addition to the expected breath sounds mentioned in section 2.4.1. Adventitious sounds are those that are superimposed over normal breath sounds. These adventitious sounds are characterized by a dominant frequency, usually over 100 Hz, and duration of more than 100 ms [78]. As early as 1957, Robertson and Coope [22] proposed a simplified classification of adventitious lung sounds into two main categories; continuous and interrupted sounds. Continuous sounds were further classified into high-pitched and low-pitched wheezes, and the interrupted sounds were divided into three categories: Coarse, medium, and fine crackles. International Lung Sound Association in 1976 further simplified the terminology: Discontinuous sound into fine and coarse crackles and continuous sound into wheeze and rhonchi [23]. Wheezes and crackles are the most widely used acoustical terms in respiratory

medicine. As this thesis is all about extracting wheezes, this section mainly focuses on explaining this particular adventitious sound. A very brief note on other adventitious sounds is also described.

1. **Wheezes:**

Wheezes are referred to as continuous musical sounds, that can last up to a whole inspiration or expiration cycle. They usually occur due to obstructions in airways when the air is being forced through small paths, which makes wheezes to give a whistling sound. These sounds occur when a collapsed airway lumen gradually opens during inspiration or gradually closes during expiration. These sounds are referred to as musical tones because as the airway lumen becomes smaller, the airflow velocity increases resulting in harmonic vibration of the airway wall and thus seems like musical tonal quality. Wheezes can be heard over the whole chest as well as the trachea, which have proven to be a good method of detecting wheezes in asthma patients [37]. The Continuity nature of wheeze helps in demonstrating a typical picture in the spectrogram. The ***Figure 2.4.4***, demonstrate the frequency spectrogram

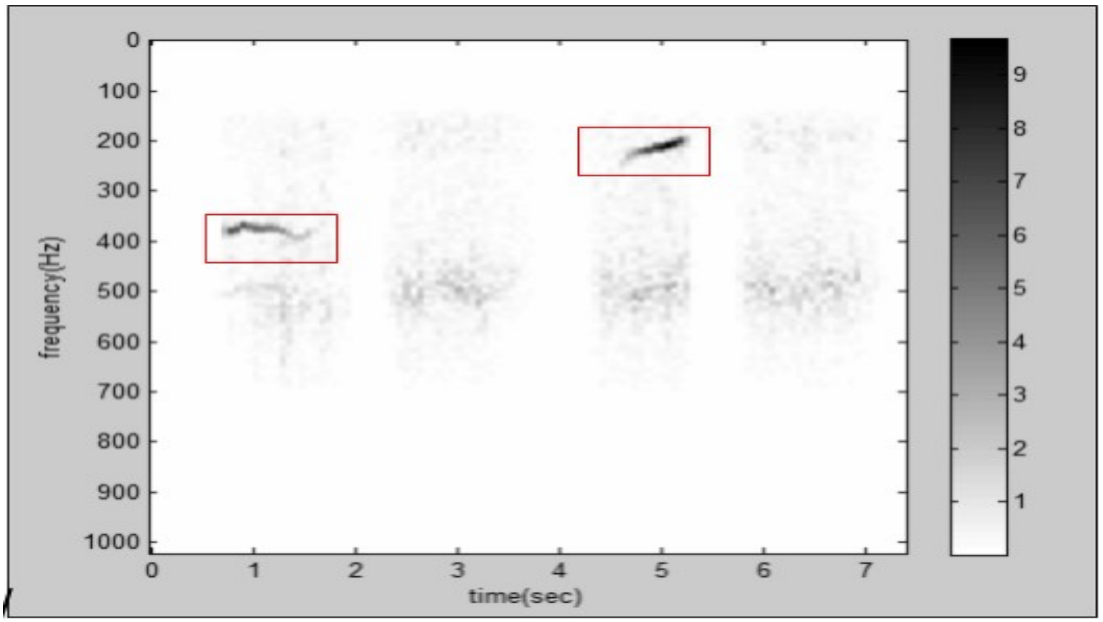


Figure 2.4.4 Spectrogram representation of respiratory signal with superimposed wheezes enclosed in red colour box

composed of wheezes (the continuous horizontal lines enclosed in red coloured box) superimposed on the breath signal.

Wheezes are referred to as “continuous” adventitious sounds because of their duration which is much longer than the “discontinuous” crackles (discussed next). As suggested by American Thoracic Society (ATS) for lung sound nomenclature the continuous adventitious sounds last for at more than 250 ms [36]. The ATS Committee also defines wheezes as high-pitched continuous sounds with a dominant frequency of 400 Hz or more [36]. According to Computerized Respiratory Sound Analysis (CORSA) continuous wheezes lie between 100 Hz and 2500 Hz with a dominant frequency between 100 Hz and 1000 Hz, and have duration greater than 100 ms [78] [85]. Therefore on a whole frequency range of wheeze sounds extends from 100 Hz to more than 1 kHz, and higher frequencies may be measured inside the airways [35]. **“**The most important point is in general Wheezes are usually louder than the underlying breath sounds”. Their intensity is higher than the normal breath sounds. From the spectrogram representation in ***Figure 2.4.4*** if we observe the spectral location of two wheezes one is at 200 Hz and the other at 400 Hz enclosed in red colour box. Also the temporal length of each wheeze is approximately 500 ms and 300 ms. Moreover from the colour bar provided in the figure we can observe that the intensity of wheezing sounds is higher than the breath sounds. All these points drawn from observing the ***Figure 2.4.4*** justify the description given on analogy of wheezes.

Wheezes vary a lot from person to person, and the sound depends on the cause, severity, and auscultation method and location. Wheezes are indications of respiratory conditions such as asthma attacks and different types of allergies that cause narrowing or obstruction of airways. Wheezing can also occur in healthy lungs when the airflow velocity increases during physical exercise. Wheezing occurs in both inspiration and expiration phases. But it is frequently found in the expiration phase. If the wheezing is in the inspiratory phase, it is an indicator of stiff stenosis whose causes range from tumours to scarring [86]. If wheezing occurs in the expiratory phase of respiration it is usually connected to the bronchiolar disease. Wheezes can be Monophonic or Polyphonic.

- Monophonic wheezing consists of a single musical note starting and ending at different times. The presence of a foreign body or accumulation of mucus or inflammation caused by bronchostenosis can produce this sound. In the case of rigid

obstruction, the wheeze is audible throughout the respiratory cycle, and when the obstruction is flexible, wheeze may be inspiratory or expiratory [38]. ***Figure 2.4.4*** represents the spectrogram representation of the respiratory sound with monophonic wheezes superimposed.

- Polyphonic wheezing consists of multiple musical notes starting and ending at the same time and is typically produced by the dynamic compression of the large, more central airways. Polyphonic wheeze is confined to the expiratory phase only [38]. ***Figure 2.4.5*** represents the spectrogram representation of the respiratory sound with polyphonic wheezes superimposed.

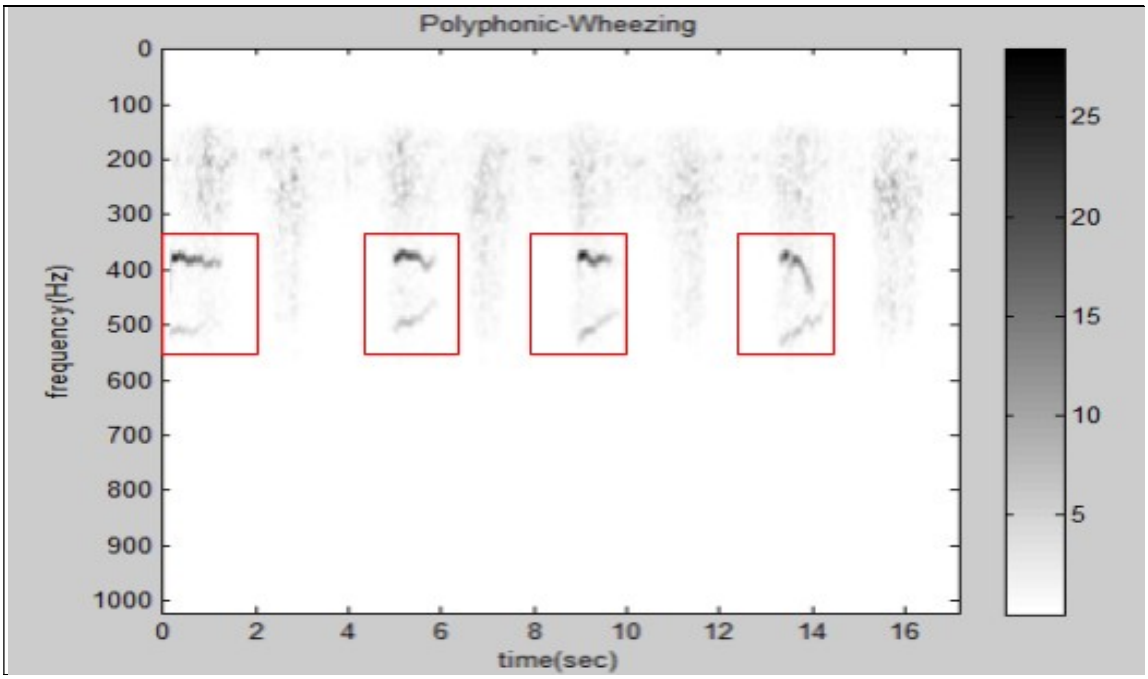


Figure 2.4.5 Spectrogram representation of a respiratory signal with superimposed "Polyphonic Wheezes"

2. Crackles

Crackles are short, explosive, and discontinuous sounds that are generally heard during the inspiratory phase. crackles usually last for shorter than 10 ms [78]. In general, crackles are caused by explosive opening of small airways [89].They are short, explosive, non-musical sounds heard on inspiration and some-times during expiration. It has a distinctly wide frequency range up to 2000Hz [88, 89]. Two types of crackles are commonly found: fine and coarse. Fine crackles are high

pitched and last for less than 5 ms as distinguished from coarse crackles, which are low pitched and have duration of about 10 ms [90].

Initially, production of crackles was attributed to the passage of air through the accumulated secretions within the large and medium-size airways, creating the bubbling sounds. However, persistence of crackles after coughing in many patients and predominant localization in inspiration argues against this theory [90] Moreover, air passing through secretions would produce both inspiratory and expiratory sounds. Later on, Forgacs proposed the second theory of crackles. According to Forgacs, small airways that were collapsed during expiration snap open during inspiration as a gradient of gas pressure is developed across the collapsed airways. Sudden explosive opening of the collapsed airways induces a rapid equalization of gas pressures resulting in oscillations of the gas column and development of crackles [91]

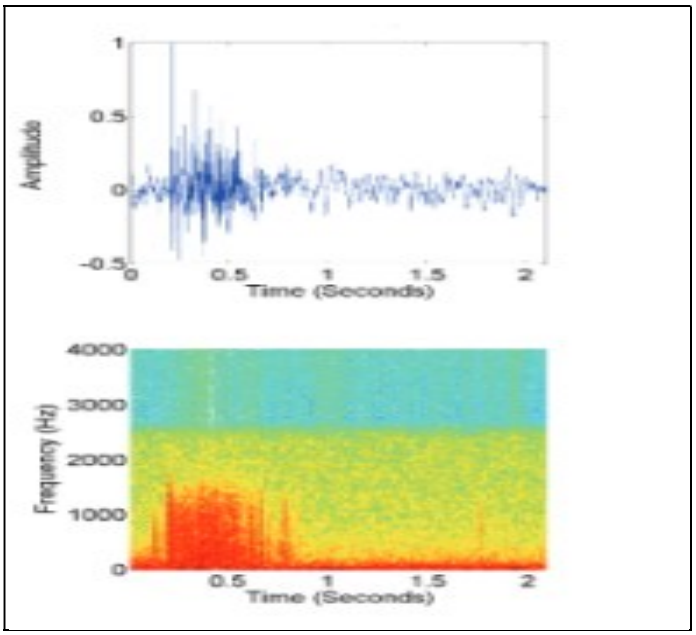


Figure 2.4.6 Time domain and frequency domain representation of breath sound with crackles [87]

2.5 State of the art methods

Variable Nomenclature

Before explaining the state of the art methods we are going to provide a clear insight of different variable used in this section

$x[n]$:- Input respiratory mixture signal composed of both respiratory and wheezing sounds.

X :- Magnitude spectrogram matrix of input respiratory mixture signal $x[n]$.

$\hat{X}$:- Estimated magnitude spectrogram matrix of input respiratory mixture signal $x[n]$.

W and H :- NMF decomposition matrices which contain frequency and time characteristics of respiratory mixture signal X

F :- Number of frequency bins

T :- Number of time bins

K :- Number of basis vectors

L :- Number of segmented blocks of respiratory mixture signal X

l :- This variable is used to represent particular segment of the L-segmented frames or blocks

$\hat{X}_R$:- Estimated magnitude spectrogram matrix of respiratory sounds

$\hat{X}_W$:- Estimated magnitude spectrogram matrix of wheezing sounds

$\hat{x}_r[n]$:- Time domain representation of estimated respiratory signal

$\hat{x}_w[n]$:- Time domain representation of estimated wheezing signal

$x_r[n]$:- Time domain representation of original respiratory signal that is mixed with input respiratory mixture signal $x[n]$

$x_w[n]$:- Time domain representation of original wheezing signal that is mixed with input respiratory mixture signal

Υ : Power Spectrogram of X

$\hat{\Upsilon}_w = \hat{\Upsilon}_h$ Power Spectrogram of estimated wheezing signal or estimated harmonic signal

$\hat{\Upsilon}_r = \hat{\Upsilon}_p$ Power Spectrogram of estimated respiratory signal or percussive signal

M : Binary mask matrix

$M_w = M_h$ Binary mask matrix for wheezing sound

$M_r = M_p$ Respiratory sound or percussive sound binary mask matrix

The main purpose of our thesis research is to extract wheezes from single-channel respiratory audio mixtures to allow physicians to easily examine the wheezes and analyze the patient's respiratory issues more keenly. This section explains how we approached the solution of NMPCF to obtain the proposed thesis statement and we are going to discuss about some of the widely used source separation approaches that we can apply for separation of wheezes from a respiratory mixture signal.

As discussed that manual auscultation process result in many drawbacks. So, to resolve the drawbacks attained through auscultation process, many researches had begun in early 1980's to recognise respiratory pathologies using computerized respiratory sound analysis. The first and foremost drawback of the auscultation process is the interference of environment sounds, heart sounds and other internal and ambient noises. Therefore most of the research studies were made to mainly suppress these interfered noises that come along with the lung sounds. The reference papers [16] [17] [18] [19] are some of the studies that are made to degrade the interfered noise in lung sounds. In [17] the authors used Savitzky-Golay (SG) filter to implement the denoising approach on lung sounds and results ended up with the best denoising performance. Whereas papers [18] [19] are studies made on suppression and cancellation of heart sounds from lung sounds through time-frequency filtering and 2D interpolation in time-frequency domain.

Whereas other important challenging tasks are the wheeze detection, classification and separation of wheezes from respiratory sounds. There are some reference studies [20-22] made on wheezing detection and obtained robust performance results. But the reference paper [20] is one of the most cited papers whose authors made research on detection of wheeze sounds based on time-frequency analysis. The experimental and testing results of this paper also justified the efficient performance and high noise robustness compared to other wheeze detection approaches. Also there are other researches like Autoregressive modelling (AR) [21] [23], wavelet transform [24] [25] [26] and Mel frequency cepstral coefficients (MFCC) that are used in feature extraction techniques to distinguish between wheeze and non wheeze pulmonary sounds [23] [27]. Similarly, a method of continuous wavelet transform that was described in [71], combined with a scale-dependent threshold seemed to provide a higher and good detection rate. Wheeze detection with wavelet

transform methods is done by decomposition of wavelet packets in two stages with first being wheeze extraction with frequency detection and then inverse transforming with reconstruction of the useful signal. Moreover there are related works done in artificial neural networks for classification and prediction of breathing sounds [23] [24] [25] [28]. The best results for wheeze and normal sounds classification were obtained with Gaussian Mixture Model (GMM) [23] [27] [28] and MFCC in 2009 [3].

However very few researches are made on separation of wheezes from respiratory mixtures. From the studies and researches done, mostly Non-Negative Matrix factorisation techniques are used for source-separation in biomedical signal processing. Moreover this technique is treated as the widely used tool to extract the target sounds from any source signal and can be used as the novel way in wheeze/respiratory sound separation.

2.5.1 Non-Negative Matrix Factorisation

Non - Negative Matrix Factorization (NMF) is a powerful matrix decomposition technique that approximates a nonnegative matrix by the product of two low-rank nonnegative matrix factors, to elaborate more specifically any non-negative matrix X is approximated by the product of two non-negative low-rank factor matrices W and H [39]. Nonnegative matrix factorization (NMF) is so popular because it automatically extracts sparse and meaningful features from a set of nonnegative data vectors, which is useful in analyzing high-dimensional data. The main difference between NMF and other factorization methods, such as SVD, is the non negativity, which assists in recovering hidden features of the input data. This is demonstrated in [42]. Unlike other matrix factorization approaches, NMF considers the fact that most types of real-world data, particularly all images, videos, audios are nonnegative and maintain such non-negativity constraints in factorization. That is why NMF is a widely used tool in image processing to learn images [40] and in audio processing for target source separation. In this particular section of the paper, we are going to explain how NMF is used in audio source separation.

NMF Algorithm:

NMF considers $x[n]$ an input mixture signal which is composed of target sources that are needed to be separated and other sources that are mixed in the original signal otherwise called non-target sources. One of the advantages of the NMF is to analyze the spectral and temporal characteristics of sounds contained in a given spectrogram. So, the input matrix is converted into time-frequency domain by using STFT (Short-time Fourier transform) If we denote X as the magnitude spectrogram of the input matrix, then each element X^{ft} represents the magnitude of the spectrum of the f^{th} frequency bin at the t^{th} time frame can be calculated by equation 2.5.1:

$$X^{ft} = \left|\sum_{n=0}^{N-1} s_t(n) e^{-j\frac{2\pi}{N}fn}\right| \qquad (2.5.1)$$

Mostly data is always available in the form of matrices. So, processing such type of data can be entailed by factorizing the input matrix X. The input matrix X is approximately factorized into two-factor matrices W and H which can be represented by equation 2.5.2. Where both W and H comprise only non-negative values.

$$X = \hat{X} = WH \qquad (2.5.2)$$

If we consider X as an F-by-T matrix with F frequency bins and T time-frames then W is considered as an F-by-K matrix where K is decided by the user which determines the number of the basis vectors likewise H is a matrix with the dimension of K-by-T. Therefore W is regarded as a basis matrix that is optimized for the linear approximation of the data in V and it contains all the frequency-related information of the active target and non-target sources. Whereas on the other side H contains the respective time-domain activations of the spectral basis in a given time-frame. In general, the value of K is chosen to be smaller than the F or T to make W and H smaller than the original matrix X. This process results in a compressed version of the original data matrix. As few basis vectors will be used for representing many data vectors, the good approximation can be achieved only if the basis vectors are capable of discovering the relevant data of the input mixture provided. ***Figure 2.5.1*** depicts the general NMF- based spectral analysis for an audio signal as explained above.

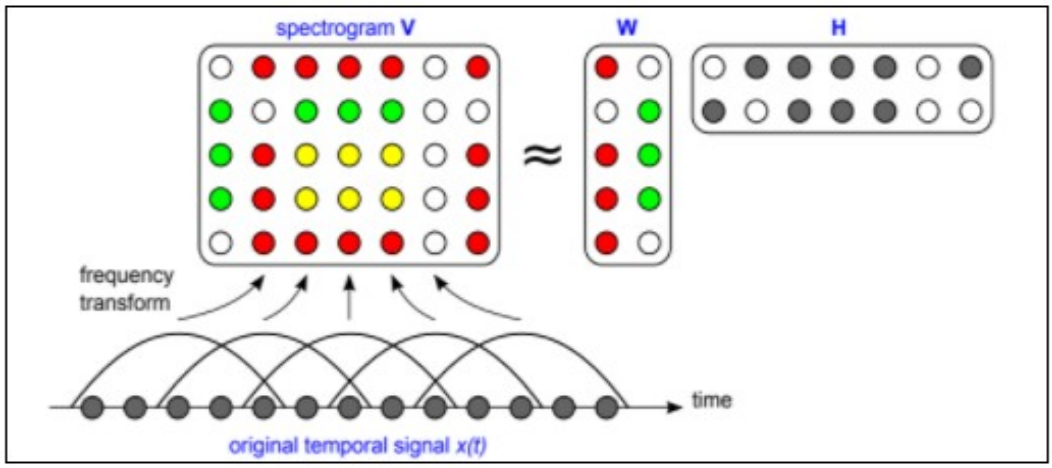


Figure 2.5.1 Pictorial representation NMF-based audio spectral analysis. A short-time frequency transform is applied to the original time-domain signal (t) . The resulting nonnegative matrix is factorized into the nonnegative matrices W and H. In this schematic example the red and green elementary spectra are unmixed and extracted into the basis matrix W. The activation matrix H returns the mixing proportions of each time-frame (a column of W). Reference [41].

The values of W and H can be estimated by solving the minimization problem as described in equation 2.5.3.

$$min_{W,H}\, D(X \mid WH) \qquad W \geq 0, H \geq 0 \qquad (2.5.3)$$

The optimization problem can be solved by iterating the matrices multiple times by using the multiplicative update rules. This helps in reducing the cost function $d(X \mid \hat{X})$, which is measured by finding the distance between X and $\hat{X}$ (X and WH). One of the most used measures can be done by simply finding the square of the Euclidean distance between X and the product of W and H.

$$d_{euc}(X \mid \hat{X}) = \|X - WH\|^2 = \sum_{i,j} \left(X_{i,j} - (WH_{i,j})\right)^2 \qquad (2.5.4)$$

This is lower bounded by zero and converges to zero if and only if $X = \hat{X}$ (where $X = WH$). This is also called the Frobenius form. But there are two other cost functions that are widely used in the field of audio processing: α-divergence and β-divergence function. Between these two, β - divergence function has a huge success rate in the field of audio signal processing. The general equation for β - divergence can be defined as :

$$d_\beta\left(X|\hat{X}\right) = \frac{1}{\beta(\beta-1)}\left(X^\beta + (\beta-1)\hat{X}^\beta - \beta X\hat{X}^{\beta-1}\right), \beta \in R\backslash\{0,1\} \quad (2.5.5)$$

The two special cases: when β = 0 equation 2.5.5 reduces to Itakura-Saito (IS) equation 2.5.7 and when $\beta = 1$ equation 2.5.5 reduces to generalized Kullback-Leibler (KL) divergence equation 2.5.6. Whereas $\beta = 2$ corresponds to the quadratic cost function, this is not widely used while dealing with non-negative data because it generates negative values sometimes.

$$d_{KL}\left(X|\hat{X}\right) = X\log\frac{X}{\hat{X}} - X + \hat{X} \qquad \beta = 1 \qquad (2.5.6)$$

$$d_{KL}\left(X|\hat{X}\right) = \frac{X}{\hat{X}} - \log\frac{X}{\hat{X}} - 1 \qquad \beta = 0 \qquad (2.5.7)$$

The β-divergence has a noteworthy property w.r.t scaling of the data. equation 2.5.8 holds for any value of β.

$$d_\beta(\lambda_X|\lambda_{\hat{X}}) = \lambda^\beta d_\beta\left(\frac{X}{\hat{X}}\right) \qquad (2.5.8)$$

The IS-divergence is the only one that shows scale-invariance property among others $d_\beta(\lambda_X|\lambda_{\hat{X}}) = d_\beta\left(\frac{X}{\hat{X}}\right)$. The Kullback-Leibler cost function is more widely applied than the Itakura-Saito function because of its scale-invariance property as discussed above. IS function depends heavily on scaling the matrices and learning them while factorising requires many iterations [44]. In the paper [43] NMF with Itakura- Saito is explained.

By considering original spectrogram matrix X and approximated spectrogram matrix $\hat{X} = WH$, equation 2.5.6 can be expressed by KL-divergence as follows:

$$d_{KL}(X|WH) = \sum_{i,j}\left(X_{i,j}\log\frac{X_{i,j}}{WH_{i,j}} - X_{i,j} + (WH)_{i,j}\right) \qquad (2.5.9)$$

The values of W and H are initialized with random values and they are iterated for several iterations using multiplicative update rules derived by minimizing the cost function equation. The multiplicative update rules derived for matrices W and H mentioned in equation 2.5.2 by minimizing equation 2.5.4 (the Euclidean cost function) are given as follows:

$$W \leftarrow W \otimes \frac{XH^T}{WHH^T} \quad (2.5.10)$$

$$H \leftarrow H \otimes \frac{W^T X}{W^T WH} \qquad (2.5.11)$$

If we consider A and B as two variables then $A \otimes B$ and $\frac{A}{B}$ corresponds to element-wise multiplication and division. Whereas equation 2.5.8 can be minimized as follows :

$$W \leftarrow W \otimes \frac{\frac{X}{WH} H^T}{1H^T} \qquad (2.5.12)$$

$$H \leftarrow H \otimes \frac{W^T \frac{X}{WH}}{W^T 1} \qquad (2.5.13)$$

From the equations 2.5.12 and 2.5.13, '1' corresponds to F-by-T matrix containing only one's. Its dimension is same as that to matrix X. Multiplicative update rules are used for updating the values of matrices W and H to attain the best quality of approximation. This implies running the algorithm for a repetitive number of iterations guarantees that the matrix values locally converge to optimal matrix factorization.

KL divergence cost function is widely used in the sound separation processes. As per the description given in [44] the factorizations obtained for $\beta > 0$ (like quadratic and KL divergence cost function) depend mostly on large data values and they are less precise while estimating lower power components. Contrastingly for factorizations obtained when $\beta \leq 0$ (like IS divergence) depends mostly on small data values and they are less precise while estimating higher power components. So choosing the cost function for deriving multiplicative update rules plays an important role while dealing with different types of data and the necessity of the data that we need to retrieve.

In our proposed thesis study of wheezing sound source separation, KL-divergence is widely applied for audio source separation. In the paper [44], the authors compare and throw good reasons why IS-divergence is replaced with KL-divergence to obtain better separation results.

Basically sound source separation is done by using supervised learning or unsupervised learning. Unsupervised source separation involves the separation of the target source from the mixture signal without having prior information and an active number of sources. On a brief note, both the basis and its activation matrices are estimated from the signal to be separated. That is why this method is also termed as Blind Source Separation (BSS) method. As far as the explanation given for NMF it

comes under unsupervised approach and gives a basic idea of how target source can be separated using matrix factorisation technique.

The main drawbacks of NMF are

i) Sometimes the matrices won't reach their global local minimum, which leads to poor convergence of the signal [50] resulting in poor separation performance.

ii) They have difficulty in clustering the decomposed spectral patterns into a specific target sound.

With the advancements in the mathematical models, the NMF methods have been further researched by adding additional parameters to improve the separation efficiency. To solve the aforementioned drawbacks, Supervised NMF (SNMF), Semi-Supervised NMF (SSNMF) [45] [46]. and Constrained NMF (CNMF) approaches has been proposed. SNMF for sound source separation is done in two stages: training stage and separation stage. In the training stage, the basis matrices of both the target source and non-target source are obtained in advance by the training signals or prior knowledge of both the sources. This step is followed by the separation stage where the basis matrices are kept fixed and activation matrices are updated for every iteration. This results in the separation of the target source components and non-target source components. The other type of sound source separation approach called SSNMF [49] is also done in two-stages: training stage and separation stage. But the main difference between SSNMF and SNMF approach is that in the training process of SSNMF approach, training signal of only target source is considered. Therefore in the training process the basis matrices of target source are collected and in the separation process they are fixed and the basis matrix of non target source and the activation matrices of both the sources are iterated to update the values for better approximation. Wheezing Sound Separation Based on Constrained Non-Negative Matrix Factorization (CNMF) [30] is one of those few papers that aimed at separation of wheezes from respiratory sounds. In this research a mixture signal, composed of wheezes and respiratory sounds, is decomposed using a cost function to integrate spectral features, spectral smoothness and sparseness, into the NMF decomposition process. These features model the spectral behaviour of wheezes applying sparseness in frequency (spectral peaks) and the respiratory sounds are modelled by applying smoothness in time (for magnitudes that vary slowly in time) and smoothness in frequency (the energy

slowly is reduced in frequency). As this approach does not require any training signal to analyse the results, the proposed method is an unsupervised (blind) method.

Also to deal with the problem of clustering of basis vectors in order to separate the target sources there are approaches like Shifted NMF [51] [52] and NMF based on spectro-temporal clustering to extract heart sounds [97]. Shifted NMF was proposed to avoid the problem of clustering, provided that a log frequency resolution is used for the frequency basis functions.

Along with NMF there are few researches which are widely used in combination with other techniques or used solely to separate the target source from mixture signal. Among these we are going to explain few which can be widely used in this research field.

2.5.2 Spectral Subtraction

The spectral subtraction method is a widely used conventional method for single channel speech enhancement to reduce the noise in the signal. But by the working principle of spectral subtraction method we have the possibility of separating wheezes from respiratory mixture signal. According to the principle of spectral subtraction, it considers a mixture signal (composed of target and non-target sources) where the target source has to be non-stationary and time-variant. Whereas the non-target source has to be stationary and the mixture signal is processed frame-by frame [93]. Respiratory sounds are non-stationary because of the change in lung volume and wheezing sounds are stationary by their characteristics [92]. All the equations and explanations given in this section are derived by taking the reference paper [93] as a guide. If we consider respiratory mixture signal $x[n]$ composed of pure respiratory sound $x_r[n]$ and wheezing sounds $x_w[n]$ which is represented in equation

$$x[n] = x_r[n] + x_w[n] \qquad (2.5.14)$$

As discussed the spectral subtraction process is done frame-by frame the respiratory mixture signal is segmented into L time frames of respiratory events and the equation is represented in its Short Time Fourier Transform as

$$X(\boldsymbol{w}, l) = X_R(\boldsymbol{w}, l) + X_W(\boldsymbol{w}, l) \qquad (2.5.15)$$

Where l represents the l^{th} segmented frame of the respiratory mixture signal. To make the equation look more simple we drop k to express the further equations. Equation 2.5.15 is expressed in its power spectrum form as

$$|X(\boldsymbol{w})|^2 = |X_R(\boldsymbol{w})|^2 + |X_W(\boldsymbol{w})|^2 \qquad (2.5.16)$$

The respiratory signal is estimated by subtracting the estimated wheezing sounds from the received spectrogram which is represented in equation 2.5.17

$$|\hat{X}_R(\boldsymbol{w})|^2 = |X(\boldsymbol{w})|^2 - |\hat{X}_W(\boldsymbol{w})|^2 \qquad (2.5.17)$$

$\hat{X}_R$ and $\hat{X}_W$ represents the estimated respiratory and wheezing signals. The estimated wheezing signal is obtained by averaging all the time frames of the respiratory mixture signal as follows:

$$|\hat{X}_w(\boldsymbol{w})|^2 = \frac{1}{L}\sum_{j=0}^{L-1}|X_j|^2 \qquad (2.5.18)$$

With the stationary characteristic feature of the wheezing sound the estimated wheezing signal from equation 2.5.18 converges to optimal power spectrum estimate as longer average is taken. The estimated respiratory signal and wheezing signal are recovered in the time-domain by using the Inverse Fourier Transform (IFT).

The spectral subtraction method, although paves a path to separate the wheezing sounds and respiratory sounds, it has some severe drawbacks. 1) From (2.5.18), it is clear that the effectiveness of spectral subtraction is heavily dependent on accurate wheezing estimation, which is a difficult task to achieve in most conditions. 2) When the wheeze estimate is less than perfect, two major problems occur, remnant wheeze in respiratory signal and poor separation.

2.5.3 Harmonic Percussive Source Separation (HPSS)

HPSS is one of the widely used audio processing techniques applied to source separation approaches. The goal of HPSS approach is to decompose a given input mixture signal into a sum of two component signals, one consisting of all harmonic sounds and the other consisting of all percussive sounds. We followed reference paper [96] as a guide to explain about this approach. The harmonic and percussive sounds have different features that can be clearly observed in their spectrogram representation which allows us in differentiating the harmonic sounds and percussive sounds. Also, the core observation in many HPSS

algorithms is that in a spectrogram representation of the input signal, harmonic sounds tend to form horizontal structures (in time-direction), while percussive sounds form vertical structures (in frequency-direction). An important characteristic of percussive sounds is that they do not have a pitch but a very clear localization in time. For an example, have a look at ***Figure 2.5.2*** where you can see the power spectrograms of two signals. ***Figure 2.5.2 (a)*** shows the power spectrogram of a sine-tone with a frequency of 4000 Hz and duration of one second. This tone is as harmonic as a sound can be. The power spectrogram shows just one horizontal line. Contrarily, the power spectrogram shown in ***Figure 2.5.2 (b)*** shows just one vertical line. It is the spectrogram of a signal which is zero everywhere, except for the sample at 0.5 seconds where it is one.

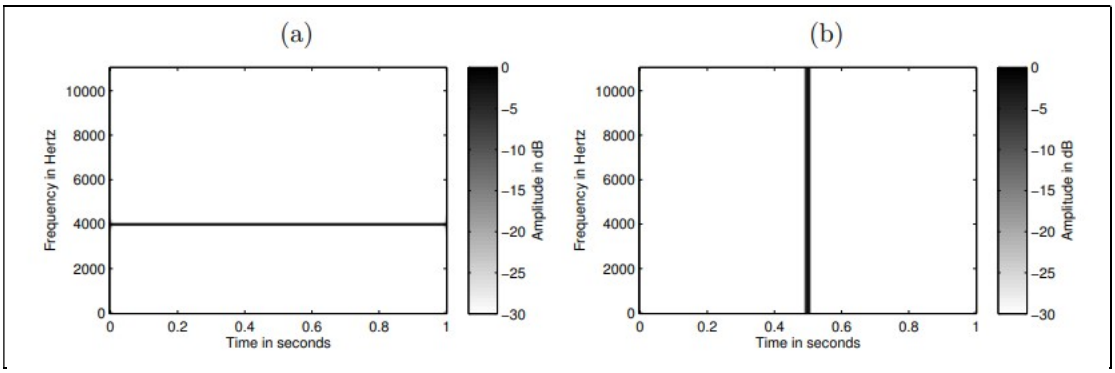


Figure 2.5.2 (a) : Spectrogram of an ideal harmonic signal (b): Spectrogram of an ideal percussive signal.[96]

However the same kind of features can be observed in ***Figure 2.4.4*** where respiratory sounds and wheezing sounds can be characterised as percussive and harmonic sounds. The separation of sources by HPSS is achieved by following four steps: 1) STFT representation of input mixture signal. 2) Median filtering, 3) Spectral masking. 4) The inversion of the STFT

1. STFT representation of input mixture signal

From above given explanation it can be stated that the harmonic and percussive sounds in a input mixture signal can be decided by their time-frequency representation. This can be achieved by applying STFT to input mixture signal. If $x[n]$ represents the input respiratory mixture signal (composed of harmonic wheezing sounds and percussive respiratory sounds) then by applying STFT to $x[n]$, the magnitude spectrogram of the input mixture signal is represented as X, then each

element X^{ft} represents the magnitude of the spectrum of the f^{th} frequency bin at the t^{th} time frame can be calculated by equation :

$$X^{ft} = \left|\sum_{n=0}^{N-1} s_t(n) e^{-j\frac{2\pi}{N}fn}\right| \qquad (2.5.3.1)$$

From X the power spectrogram of $x[n]$ can be derived by using equation (2.5.3.2) which is represented by Υ

$$\Upsilon = |X|^2 \qquad (2.5.3.2)$$

2. Median filtering

The next step in HPSS process is to compute a harmonically enhanced spectrogram $\widehat{\Upsilon}_h$ and a percussively enhanced spectrogram $\widehat{\Upsilon}_p$by filtering Υ. This is achieved by applying median filtering to input mixture's power spectrogram signal Υ. Median of set of numbers is found by arranging all numbers from lowest to highest value and picking the middle one. For example if we let $A = \{a_n \epsilon\, \mathbb{R} | n\, \epsilon\, [0: N-1]\}$ be a set of real numbers of size N. Then, the median of A is defined as:

$$median(A) \coloneqq \left\{ \begin{array}{ll} a_{\frac{N-1}{2}} & for\ N\ being\ odd \\ \frac{1}{2}\left(a_{\frac{N}{2}} + a_{\frac{N}{2}+1}\right) & for\ otherwise \end{array} \right\}$$

Therefore for a given matrix $\Upsilon \in \mathbb{R}^{F\times K}$ the enhanced spectrograms are computed by harmonic and percussive median filters as

$$\widehat{\Upsilon}_w = \widehat{\Upsilon}_h = medfilt_h(\Upsilon) \coloneqq median(\{\Upsilon(m - l_h, k), \dots, \Upsilon(m + l_h, k)\})$$

$$\widehat{\Upsilon}_r = \widehat{\Upsilon}_p = medfilt_p(\Upsilon) \coloneqq median(\{\Upsilon(m, k - l_p), \dots, \Upsilon(m, k + l_p)\})$$

Where $F, K, l_p, l_h \in \mathbb{N}$, where $2l_h + 1$ and $2l_p + 1$ are lengths of the median filters respectively where $m\, \epsilon\, [0: F-1]$ and $k\, \epsilon\, [0: K-1]$ for$\Upsilon \neq 0$.

3. Spectral masking

Next step is to assign the time-frequency bins of X to enhanced harmonic($\widehat{\Upsilon}_w = \widehat{\Upsilon}_h$) and percussive ($\widehat{\Upsilon}_r = \widehat{\Upsilon}_p$) power spectrograms which is achieved by applying binary masking. A binary mask is a matrix $M\epsilon\{0,1\}^{F\times K}$which is applied on to X by calculating $X \odot M$. A mask value of 1 preserves the value in the STFT and 0 suppresses it. In HPSS algorithm the masks are decide by comparing the values in $\widehat{\Upsilon}_h$ and $\widehat{\Upsilon}_p$.

$$M_w(m,k) = M_h(m,k) \coloneqq \left\{ \begin{matrix} 1 & if\ \hat{Y}_h(m,k) \geq \hat{Y}_p(m,k) \\ 0 & else \end{matrix} \right\} \quad (2.5.3.3)$$

$$M_r(m,k) = M_p(m,k) \coloneqq \left\{ \begin{matrix} 1 & if\ \hat{Y}_p(m,k) > \hat{Y}_h(m,k) \\ 0 & else \end{matrix} \right\} \quad (2.5.3.4)$$

Having the wheezing and respiratory masks $M_w(m,k)$ and $M_r(m,k)$ from equation (2.5.3.3) and equation (2.5.3.4) when applied on to the original spectrogram of input mixture signal X yields the spectrograms of the harmonic (wheeze) and percussive (respiratory) components of the signal $\hat{X}_W \coloneqq (X \odot M_w)$ and $\hat{X}_R \coloneqq (X \odot M_r)$. There is important point to note that by the definition $M_h(m,k)$ and $M_p(m,k)$ holds that $M_w(m,k) + M_r(m,k) = 1$ for $m \in [0: F-1]$ and $k \in [0: K-1]$. Every time-frequency bin of X is either assigned to $\hat{X}_W$ or $\hat{X}_R$

4. Inversion of STFT

By applying inverse STFT to the estimated spectrograms $\hat{X}_W$ and $\hat{X}_R$ obtained from step3 we obtain time domain signals of separated wheezing sounds $\hat{x}_w[n]$ and respiratory sounds $\hat{x}_r[n]$.

From the HPSS approach we came to know about two different techniques involved i.e. spectral filtering and spectral masking which are useful for separating sources from mixture signal. There are several researches done on audio source separation using spectral filtering and spectral masking. With the working idea of HPSS there were few researches made on sound source separation techniques which used NMF and HPSS to obtain better separation results. [98] [99] are few of those reference papers.

As explained above many researches were done on sound source separation but with the introduction of matrix co-factorisation techniques in source separation field has greatly improved the separation performance. In general Matrix co-factorisation methods perform two (or more) factorisations in parallel, which are linked in a particular way. In particular, these methods are well suited to music signal analysis. There are few researches done on source separation approaches using Non-Negative matrix Co-Factorisation(NMCF). [100] [101] are some of the refrence studies that give information on NMCF applied to music signal analysis and soft NMCF. The use of NMCF methods increase the separation efficiency of the estimated signals.

Advancements in NMCF approaches led to the development of Non-Negative Matrix Partial Co-Factorisation (NMPCF) approaches that tends to separate most of the interfered sources with the target source than the NMF and NMCF approaches. Recently a study was done on Wheezing Sound Separation Based on Informed Inter-Segment Non-Negative Matrix Partial Co-Factorization [102] in which the authors incorporated a new feature of inter-segment information, informed by a wheezing detection system, into the factorization to reconstruct a more accurate modelling of normal respiratory sound and their results also demonstrated the significant improvement obtained in the wheezing sound quality compared to other approaches.

In our thesis study we work on NMPCF approaches to segregate the wheezing sounds from respiratory audio mixtures and we explain about this clearly in Materials and Methods chapter 4.

3 OBJECTIVES

1. Compilation and study of state-of-the art references related to main topic of this Master Thesis
2. Design and implementation of proposed algorithm in MATLAB in order to correctly separate wheezing sounds from respiratory audio mixtures.
3. Creation of database composed of wheezes and respiratory sounds from available internet pulmonary repositories
4. Evaluation of the implemented system.
5. Development of user-friendly interface in MATLAB for the user.

4 MATERIALS AND METHODS

In this particular section of the draft we will discuss how we approached to the solution of the proposed methodology. To assess and evaluate the performance of the proposed methodology a dynamic program has been developed in MATLAB as mentioned in the objectives section 3. This program is utilised to segregate wheezes and normal respiratory sounds from the respiratory audio mixture signal. A comprehensive evaluation has been carried out by taking into account different characteristics of respiratory sounds. The review is focused mainly on studies based on partial matrix co-factorisation approaches that are carried out to separate wheezes and normal respiratory events. As already discussed in section 2.5, NMF is one of the robust algorithms used for sound source separation when sources and components of signal are dependent, but under the conditions that additional constraints are imposed, like sparsity, non-negativity or smoothness. However, without any prior knowledge of a target source signal, the standard NMF cannot separate specific source signals from the mixture signal. To tackle this problem, recent advances emerged a new concept of joint decomposition or collective matrix factorization called Non-Negative Matrix Partial Co-Factorization (NMPCF). In this approach instead of following a simple initialization process it partially co-factorises the signal. The NMPCF process for source separation is explained in three different approaches.

1. NMPCF for spectral source separation (S-NMPCF) [31]
2. NMPCF for temporal source separation (T-NMPCF) [32]
3. NMPCF for spectral and temporal source separation (ST-NMPCF). [33]

The third approach, NMPCF for spectral and temporal source separation is our proposed methodology which proves our thesis statement. In this section a brief explanation on the first two approaches of NMPCF is given to the reader and in the third approach we will clearly explain how these two approaches combined gave rise to a new technique that improves the separation efficiency. Below we are going to provide nomenclature of all the necessary variables that we are going to use in this whole section to make it easy for the reader to understand them properly.

Variable Nomenclature

$x[n]$:- Input respiratory mixture signal composed of both respiratory and wheezing sounds

$y[n]$:- Respiratory training signal composed of solo-respiratory sounds

X :- Magnitude spectrogram matrix of input respiratory mixture signal $x[n]$

Y :- Magnitude spectrogram matrix of respiratory training signal $y[n]$

L :- Number of segmented blocks of respiratory mixture signal X

l :- This variable is used to represent particular segment of the L-segmented frames or blocks

$\hat{X}_R$:- Estimated magnitude spectrogram matrix of respiratory sounds

$\hat{X}_W$:- Estimated magnitude spectrogram matrix of wheezing sounds

$\hat{x}_r[n]$:- Time domain representation of estimated respiratory signal

$\hat{x}_w[n]$:- Time domain representation of estimated wheezing signal

$x_r[n]$:- Time domain representation of original respiratory signal that is mixed with input respiratory mixture signal $x[n]$

$x_w[n]$:- Time domain representation of original wheezing signal that is mixed with input respiratory mixture signal

4.1 NMPCF for Spectral Source Separation (S-NMPCF)

NMPCF for spectral source separation is a supervised factorisation technique, i.e. this approach utilises training data or side information to deal with the segregation process. S-NMPCF takes the advantage of joint co-factorisation of the training data and the mixture signal i.e. the common basis vectors of the factor matrices of the training data are used by the factorised matrices of the mixture signal. If $y[n]$ represents the respiratory training signal or prior side-information signal which is composed of solo-respiratory sounds and $x[n]$ represents the respiratory audio mixture signal, then the signals $x[n]$ and $y[n]$ can be represented as follows:

$$x[n] = x_r[n] + x_w[n] \tag{4.1.1}$$

$$y[n] = y_r[n] \quad (4.1.2)$$

Note : All the formulas used in this particular section of the draft are obtained by using [31] as a reference guide

Considering respiratory audio mixture signal implies a signal composed of respiratory sounds and wheezing sounds, therefore $x[n]$ in (4.1.1) is expressed as sum of respiratory sounds $x_r[n]$ and wheezing sounds $x_w[n]$. Whereas the respiratory training signal implies signal composed of concatenation of different types of solo-respiratory sounds. Therefore the respiratory training signal $y[n]$ in (4.1.2) is expressed as $y_r[n]$. S-NMPCF uses a decomposition model which explicitly distinguishes the respiratory sounds and the wheezing sounds from the respiratory audio mixture signal as follows:

$$X \cong \hat{X} = \hat{X}_R + \hat{X}_W = A_R S_R^{(T)} + A_W S_W^{(T)} \quad (4.1.3)$$

$$Y \cong \hat{Y} = A_R V_R^T \quad (4.1.4)$$

X and Y from equations (4.1.3) and (4.1.4) represents the magnitude spectrogram matrices of respiratory audio mixture signal $x[n]$ and respiratory training signal $y[n]$ obtained by using equation (2.5.1). From matrix factorisation approaches we tend to estimate the spectrogram matrices of $x[n]$ and $y[n]$ by using the required algorithms. Therefore $\hat{X}$ and $\hat{Y}$ represents the estimated magnitude spectrogram matrices of X and Y. Whereas A_R and S_R represents the frequency and time domain characteristics of the respiratory sounds in a respiratory audio mixture signal, respectively, and A_W and S_W represents the frequency and time-domain characteristics of wheezing sounds in a respiratory audio mixture signal. All the matrices $A_R, A_W, S_R \; and \; S_W$ are non-negative. However A_R and V_R from equation (4.1.4) represents the frequency and time domain characteristics of represents the respiratory sounds in respiratory training signal and are constrained to be non-negative. Since the training signal contains only respiratory sounds the decomposition model in (4.1.4) contains matrices with frequency and time domain characteristics of solo-respiratory sounds. A_R is treated same for decomposition models in equation (4.1.3) and (4.1.4) and V_R represents the time-domain activations of each column of A_R in Y. In general the straight forward way to use the prior knowledge matrix to discriminate the respiratory sounds and wheezing sounds from respiratory audio mixture signal is to make use of the values of matrix A_R obtained by decomposing the magnitude spectrogram matrix of the respiratory

training signal Y (4.1.4) to initialise them as initial values of A_R in decomposition model (4.1.3) [94]. However, this kind of initialization approach does not guarantee that the initialized part remains to represent respiratory sounds after the decomposition process. Therefore in this approach instead of the simple initializing the values, a partial co-factorization method is implemented which shares the frequency basis matrix A_R to factorize the input respiratory audio mixture signal X and the respiratory training signal Y. The proper segregation of the respiratory sounds and wheezing sounds from respiratory audio mixture signal is done by assigning certain number of basis vectors to the respiratory and wheezing sounds. The basis vectors are defined while decomposing the matrices X and Y. K_R and K_W are considered as variables to indicate the basis vectors of respiratory and wheezing sounds. This implies A_R is of size $F \times K_R$ and it is used to reconstruct the respiratory sounds from the respiratory audio mixture signal and A_W is of size $F \times K_W$ and it is used to reconstruct the wheezing sound from the mixture signal. After decomposing the input signals X and Y into common and individual matrices, they are jointly factorized and updated for several iterations. In this process by commonly sharing the same basis vectors, $A_R^{F \times K_R}$ utilizes the spectral information of solo-respiratory sounds from respiratory training signal. As a result the first part of the equation (4.1.3) $A_R S_R^{(T)}$ partly reconstructs all the respiratory sources that are mixed in X. Contrarily the individual basis vectors $A_W S_W^{(T)}$ collect the remaining sources of input mixture (in our case they are wheezes) and partly reconstruct them, thus resulting in separated respiratory and wheezing sounds. The whole decomposition process used for segregating respiratory sounds and wheezing sounds from respiratory mixture signal is depicted pictorially in ***Figure 4.1.1***

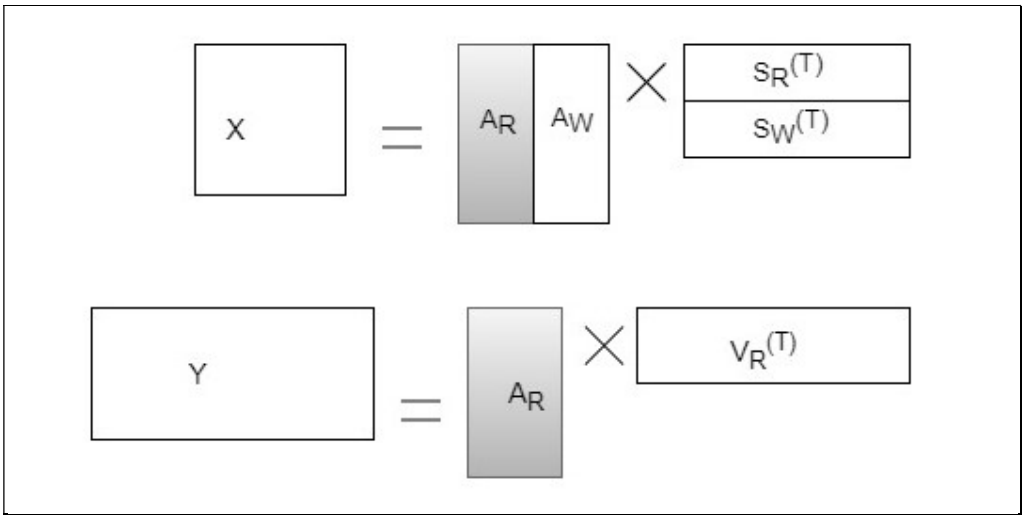


Figure 4.1.1 Pictorial representation of the S-NMPCF separation model with* X *and* Y *representing the Respiratory Mixture Signal and Respiratory Training Data respectively.

The co-factorization process is possible only with the factorization models and update rules derived. So, to implement the factorization process the input matrices have to solve the minimization problem. The concept of minimization problem is discussed in the state of the art methods while explaining the concept of NMF. Equation (4.1.3) details the objective function used by the S-NMPCF approach to minimize the residuals of the models represented in equation (4.1.3) and equation (4.1.4).

$$L = \frac{1}{2}\|X - A_R S_R^T - A_W S_W^T\|_F^2 + \frac{\lambda}{2}\|Y - A_R V_R^T\|_F^2 \qquad (4.1.5)$$

From equation (4.1.5) λ (lambda) is the parameter that is used for adjusting the relative importance between the input respiratory mixture signal and respiratory training data. This is the most important parameter whose variation leads to different outputs thus ending up with the more efficient output. To find the local minima, the gradient of the objective function is calculated. The multiplicative update rules are derived by decomposing the gradient of the objective function from equation (4.1.5). The derived multiplicative update rules should not loose their non-negativity feature during the iteration process. Therefore these are derived in such a way that they have a stationary point at the local minimum, thus not allowing the matrices to lose their non-negativity constraint**.** The gradient of the objective function (equation 4.1.5) is decomposed as follows:

$$\frac{\partial L}{\partial A_R} = \left[\frac{\partial L}{\partial A_R}\right]^{+} - \left[\frac{\partial L}{\partial A_R}\right]^{-} \quad (4.1.6)$$

From equation (4.1.6) multiplicative update rules for each matrix involved in equations (4.1.3) and (4.1.4) are derived as follows:

$$A_R \leftarrow A_R \odot \frac{\left[\frac{\mathrm{dL}}{\mathrm{dA_R}}\right]^{-}}{\left[\frac{\mathrm{dL}}{\mathrm{dA_R}}\right]^{+}} \quad (4.1.7)$$

The below equations show the gradients and multiplicative update rules derived.

The calculated gradients of the matrices are as follows:

$$\frac{\partial L}{\partial A_R} = -XS_R + A_R S_R^{(T)} S_R + A_W S_W^{(T)} S_R - \lambda \mathrm{Y V_R} + \lambda \mathrm{A_R V_R^{(T)} V_R} \quad (4.1.8)$$

$$\frac{\partial L}{\partial S_R} = -X^{(T)} A_R + S_R A_R^{(T)} A_R + \mathrm{S_W A_W^{(T)} A_R} \quad (4.1.9)$$

$$\frac{\partial L}{\partial A_W} = -XS_W + A_W S_W^{(T)} S_W + \mathrm{A_R S_R^{(T)} S_W} \quad (4.1.10)$$

$$\frac{\partial L}{\partial S_W} = -X^{(T)} A_W + S_W A_W^{(T)} A_W + \mathrm{S_R A_R^{(T)} A_W} \quad (4.1.11)$$

$$\frac{\partial L}{\partial V_R} = -Y^{(T)} A_R + V_R A_R^{(T)} A_R \quad (4.1.12)$$

Multiplicative update rules for all the matrices

$$A_R \leftarrow A_R \odot \frac{\mathrm{XS_R + \lambda Y V_R}}{\mathrm{A_R S_R^{(T)} S_R + A_W S_W^{(T)} S_R + \lambda A_R V_R^{(T)} V_R}} \quad (4.1.13)$$

$$S_R \leftarrow S_R \odot \frac{\mathrm{X^{(T)} A_R}}{\mathrm{S_R A_R^{(T)} A_R + S_W A_W^{(T)} A_R}} \quad (4.1.14)$$

$$A_W \leftarrow A_W \odot \frac{\mathrm{XS_W}}{\mathrm{A_W S_W^{(T)} S_W + A_R S_R^{(T)} S_W}} \quad (4.1.15)$$

$$S_W \leftarrow S_W \odot \frac{\mathrm{X^{(T)} A_W}}{\mathrm{S_W A_W^{(T)} A_W + S_R A_R^{(T)} A_W}} \quad (4.1.16)$$

$$V_R \leftarrow V_R \odot \frac{\mathrm{Y^{(T)} A_R}}{\mathrm{V_R A_R^{(T)} A_R}} \quad (4.1.17)$$

Based on the update rules derived in equations (4.1.13 to 4.1.17) the matrices are iterated for several iterations which results in properly estimating the basis matrix A_R for respiratory sounds and its corresponding time activations S_R. After the iteration process, the estimated separated signals are reconstructed as follows:

$$\widehat{\mathrm{X}}_{\mathrm{R}} = \mathrm{A}_{\mathrm{R}}\mathrm{S}_{\mathrm{R}}^{(\mathrm{T})} \qquad (4.1.18)$$

$$\widehat{\mathrm{X}}_{\mathrm{W}} = \mathrm{A}_{\mathrm{W}}\mathrm{S}_{\mathrm{W}}^{(\mathrm{T})} \qquad (4.1.19)$$

Equations (4.1.16) and (4.1.17) represents the magnitude spectrogram of separated respiratory sources and separated wheezes. The signals are in frequency-domain; these are then transformed into time-domain by inverse Fourier transform. The performance results of this approach will be discussed clearly in the results section of the draft.

The main advantages of using the NMPCF learning process are:- NMPCF approaches makes common basis vectors to reflect the spectral property of the given input signal. These results in ordering the basis vectors into two groups, where each of them represents one of the two sets of sources provided. Secondly, the resulting two sets of basis vectors are more separation-friendly; they are more likely to be distinctive than those of standard NMF as explained in state of art methods.

The main drawback of S-NMPCF scheme is that the algorithm adds same importance to all the frames of the input respiratory mixture signal during multiplicative iterative process. According to the characteristics of breath signal, it has repetitive respiratory events of small variations throughout its length, but the S-NMPCF approach segregates only the respiratory spectral components from the input mixture signal using the knowledge of training signal. Which result in reflecting the temporal components to still appear in the estimated wheezing signal leading to poor segregation performance. Therefore research was made on resolving this drawback which led to the study of NMPCF for temporal source separation (T-NMPCF).

4.2 NMPCF for Temporal Source Separation (T-NMPCF)

The stability of respiratory sound characteristics of any individual influences the applicability of breath sound measurements as a clinical tool. So, it is necessary that the breath-to-breath variability is small to increase the clinical importance of breath sounds[54]. Several researches have proved that these variations are small [53]. So, this implies respiratory sounds will be repetitive with small variations after every interval of time. It is necessary that the breath sounds recorded are of at least 2sec for better response. As discussed in S-NMPCF it does not segregate the temporal components present in the signal . So, T-NMPCF is an improvisation technique which tries to collect all the repeating events (respiratory sounds) present in the signal by partially co-factorizing the mixture matrix with the condition that there will be common basis vectors that resemble those repeating sound components through all the blocks. Moreover, this approach provides an alternative solution to the S-NMPCF segregation process, without using any side-information or training data. So, this approach comes under the unsupervised factorization technique.

T-NMPCF can also be called as blind source separation approach as segregation of the target source from the mixture signal is done without using any prior knowledge or training signal. As this approach don't consider any prior-knowledge or training signal only respiratory mixtures signals are considered to test the derived T-NMPCF algorithm. It is an important note to the readers that the respiratory mixture signal $x[n]$ we considered is with an assumption that there are repetitive respiratory events throughout the block with non-repetitive wheezing sounds. Moreover according to the nature of wheezing sounds they won't usually occur in every frame which is very important during co-factorisation process to segregate the temporally repetitive respiratory components from $x[n]$ leaving the wheezing components behind. As the process utilizes a temporal property and repeatability for segregating the target source this method is supposed to be called as NMPCF for Temporal source separation (T-NMPCF).

Considering $x[n]$ as the input respiratory mixture signal the magnitude spectrogram of the input mixture signal X is segmented into l-number

$(1, 2, 3, \dots, L)$ of column blocks which share the common respiratory basis components of repetitive respiratory events throughout all the blocks. The magnitude spectrogram matrices of segmented number blocks is represented as $X^{(1)}, X^{(2)}, X^{(3)}, \dots, X^{(L)}$. So, for l^{th} input matrix T-NMPCF decomposes the input magnitude spectrogram as follows :

$$X^{(l)} = X_R^{(l)} + X_W^{(l)} = A_R S_R^{(l)} + A_W^{(l)} S_W^{(l)} \quad (4.2.1)$$

Note : All the formulas or equations used in this particular section of the draft are obtained by using [32] as a reference guide

Each segmented spectrogram matrix contain respiratory components and may or may not contain wheezing components. Therefore the each segmented spectrogram matrix $X^{(l)}$is represented as a mixture of spectrogram matrices of respiratory $X_R^{(l)}$ and wheezing $X_W^{(l)}$ signals. From equation (4.2.1) the spectrogram matrix $X_R^{(l)}$ is decomposed into two matrices A_R and $S_R^{(l)}$, which represent the frequency and time characteristics of l^{th} segmented respiratory block. Where A_R shares all the respiratory bases vectors that are temporally repetitive and commonly appear in all the segmented blocks and $S_R^{(l)}$ contains its respective time-domain activations. Similarly, the spectrogram matrix $X_W^{(l)}$ is decomposed into two matrices $A_W^{(l)}$ and $S_W^{(l)}$ where, $A_W^{(l)}$ contains the frequency characteristics of wheezing components present in the l^{th} segmented block and $S_W^{(l)}$ reflects its respective time-domain activations. The number of factorization models for this technique varies with the number of column-blocks made. All these matrices are partially co-factorised and updated for several iterations using the update rules derived from objective function. ***Figure 4.2.1*** represents the pictorial representation of factorisation models used by T-NMCF approach.

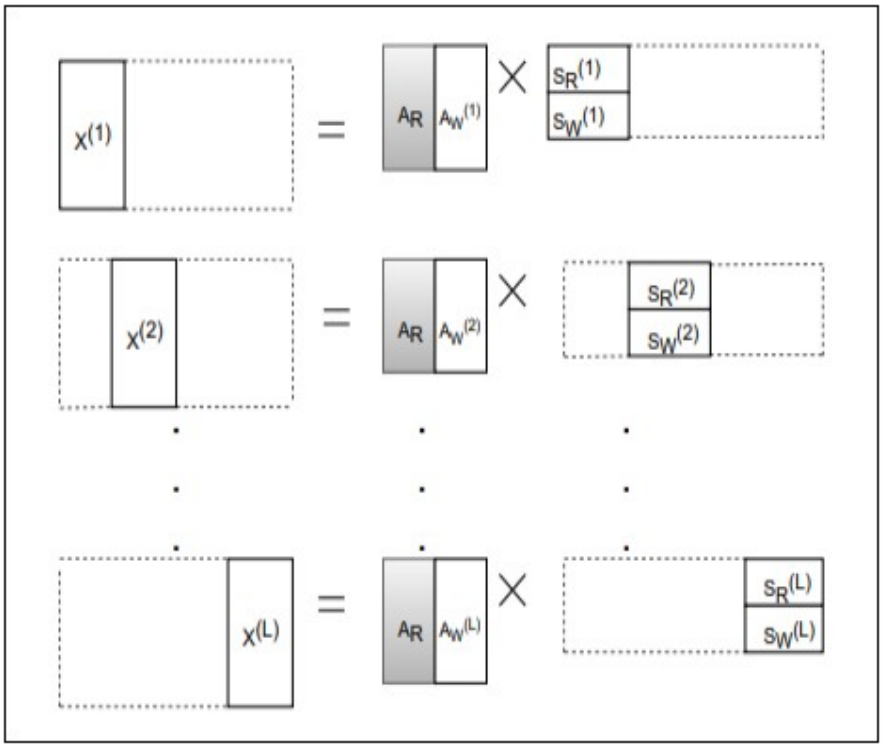


Figure 4.2.1 Pictorial representation of the decomposition model of T-NMPCF approach with $\mathbf{X^{(1)}, X^{(2)}, X^{(3)}, \ldots, X^{(L)}}$ ***representing the l-segmented blocks of input respiratory audio mixture signal*** $\mathbf{X}$

The objective function used in this approach to derive the update rules is given by :

$$\Gamma_{T-NMPCF} = \sum_{l=1}^{L} \lambda_l \left\| X^{(l)} - A_R S_R^{(l)} - A_W^{(l)} S_W^{(l)} \right\|_F^2 + \gamma \left\{ \sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2 \right\} \quad (4.2.2)$$

From equation (4.2.2) the term $\left\{ \sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2 \right\}$ is a regularisation term which is equal to $\sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2 = L \|A_R\|_F^2 + \sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2$. Regularisation is a process of adding additional information to the objective function to solve the problem of over fitting during the optimisation process. These terms help to improve the optimisation process. By using equation (4.2.2), multiplicative update rules are derived to learn the matrices $S^{(l)}$ A_R, and $A_W^{(l)}$ and which are represented in equations (4.2.3), (4.2.4) and (4.2.5) as follows

$$S^{(l)} \leftarrow S^{(l)} \odot \left(\frac{\left(A^{(l)}\right)^T X^{(l)}}{\left(A^{(l)}\right)^T A^{(l)} S^{(l)}} \right) \quad (4.2.3)$$

$$A_R \leftarrow A_R \odot \left(\frac{\sum_{l=1}^{L} \lambda X^{(l)} \left(S_R^{(l)}\right)^T}{\sum_{l=1}^{L} \lambda_l A^{(l)} S^{(l)} \left(S_R^{(l)}\right)^T + \lambda L A_R} \right) \quad (4.2.4)$$

$$A_W^{(l)} \leftarrow A_W^{(l)} \odot \left(\frac{\lambda_l X^{(l)} \left(S_W^{(l)}\right)^T}{\lambda_l A^{(l)} S^{(l)} \left(S_W^{(l)}\right)^T + \lambda A_W^{(l)}} \right) \quad (4.2.5)$$

Since, the T-NMPCF approach won't consider any axillary input training signal, the common respiratory basis vectors of all the segmented magnitude spectrogram matrices $(X^{(1)}, X^{(2)}, X^{(3)}, \ldots, X^{(L)})$ will be iterated to learn the spectral components of repetitive respiratory sounds throughout all the blocks. These segmented column-blocks are considered as input-matrices to the NMPCF algorithm. Since the input column-blocks are exclusive time periods of the mixture signal A_R contains basis vectors of only respiratory sounds after updating the matrices. Whereas $A_W^{(l)}$ contains all the non-repeatable sources in each segment. These non-repeating components cannot be shared because of its changing frequency characteristics and relatively lower repeatability in their temporal appearances. Thus, finally we end up with respiratory and wheezing sounds separated.

The drawback of T-NMPCF approach is that it segregates all the temporally repetitive components from the mixture signal i.e. if the wheezing sounds are repetitive along with the respiratory sounds in the segmented blocks, then the T-NMPCF approach segregates them also. This results in poor segregation performance. S-NMPCF and T-NMPCF approaches have their own advantages which on combined can improve the efficiency of the segregation process. Therefore our proposed approach is the combination of these two approaches which is called as NMPCF for spectral and temporal source separation (ST-NMPCF).

4.3 NMPCF for Spectral and Temporal Source Separation (ST-NMPCF)

Our proposed method is an approach which resolves flaws that occurred in previous NMPCF-based systems S-NMPCF and T-NMPCF. Our proposed scheme is a combination of previously mentioned NMPCF approaches and this is named as NMPCF for Spectral and Temporal Source Separation (ST-NMPCF) [33]. We are going to explain our

proposed methodology in four different sections to make everything clear to the reader. The four sections are as follows:

1. Overview of the proposed separation system
2. Factorisation models
3. Objective function and update rules
4. Separation procedure.

Note : All the equations and explanations used in this particular section of the draft are obtained by using [33] as a reference guide

4.3.1 Overview of the Proposed Separation System

The main aim of the ST-NMPCF approach is to resolve the flaws that occurred in S-NMPCF and T-NMPCF approaches to enhance the quality of wheezing sound by explicitly separating respiratory sounds from the mixture signal. This approach tends to enhance the quality of target source by taking the advantage of both the spectral and temporal characteristics of the respiratory signal source. ***Figure 4.3.1*** represents the flow chart diagram of our proposed methodology and we are going to explain about importance of each block in detail.

Time-Frequency conversion:
All audio signals when recorded are represented in time-domain. But we won't get much information when audio signals are represented in time-domain. They inform us only about the RMS amplitude of the signal or more specifically the average amplitude of the signal-averaged at all the various frequencies of the bandwidth and not more than that. But the frequency representation of the signal gives us information regarding the amplitude of audio components that are produced at each specific frequency. So, the frequency domain representation of the signal is necessary to view the respiratory and wheezing sound components found in the mixture signal.
From the ***Figure 4.3.1***, $y[n]$ represents the respiratory training signal which is composed of solo-respiratory sounds formed by concatenation of several pure respiratory sounds (without any abnormal sounds in it). The information regarding this signal is detailed in **section 5.1**. Whereas $y[n]$ represents the n-th sample of input respiratory mixture signal which is a composed of both respiratory and wheezing sounds. X Represents the magnitude spectrogram matrix of the input respiratory mixture signal $x[n]$, then each element of X_{ft} represents the magnitude of the spectrum

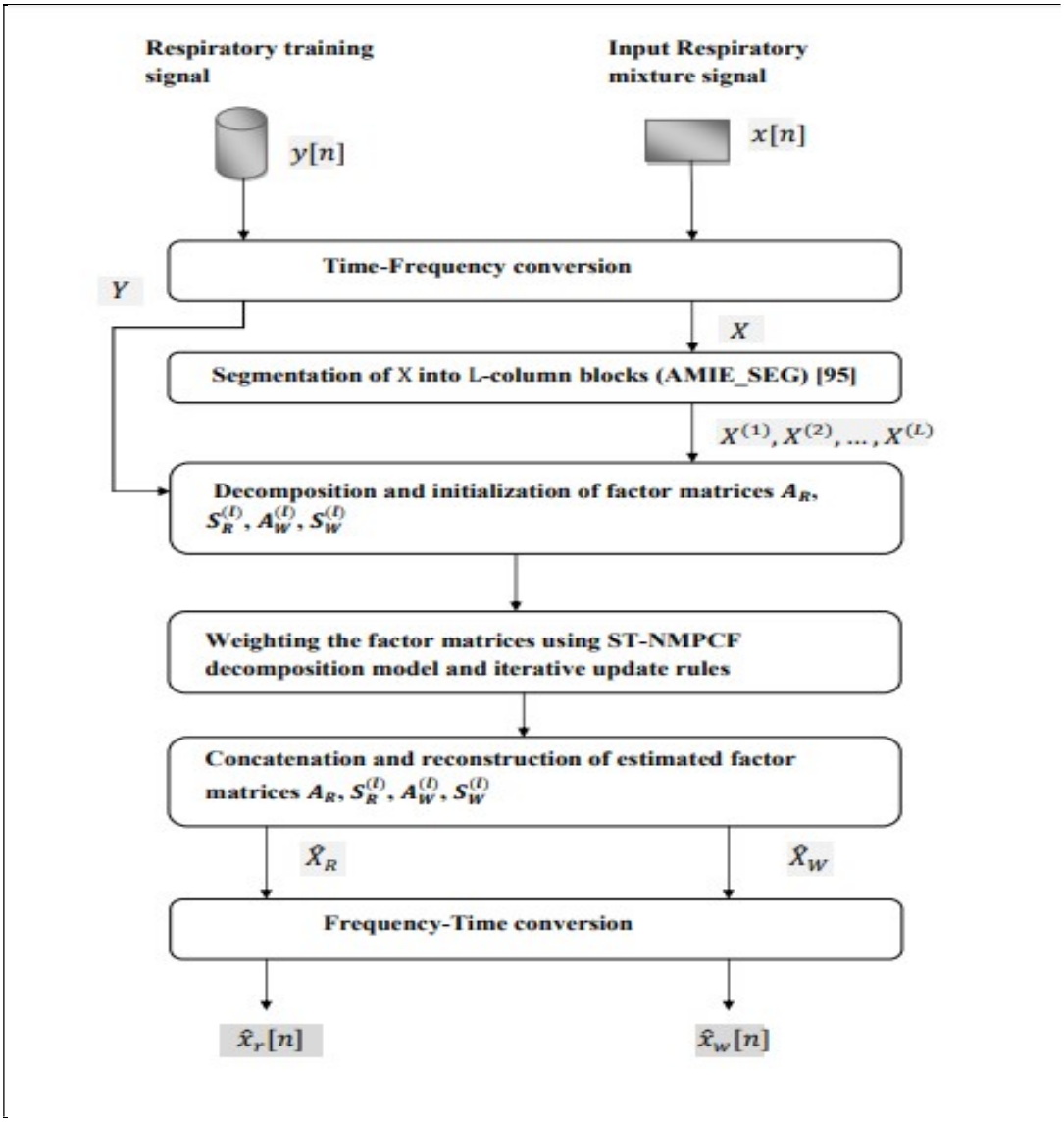


Figure 4.3.1 Flow chart diagram of the proposed ST-NMPCF approach

of the f-th frequency bin at the t-th time frame which is calculated by calculating the magnitude of Short Time Fourier Transform (STFT) as represented in equation

$$X_{f,t} = \left|\sum_{n=0}^{N-1} s_t(n) exp^{-j\frac{2\pi}{N}fn}\right| \qquad (4.3.1.1)$$

From equation (4.3.1.1) s_t represents the t-th windowed time domain signal of $x[n]$ where N number of samples are used for windowing the signal and Hamming window is used for windowing the signal with 50%

or 25% of the N-number of samples used as overlapping samples or hop-samples. As the input respiratory mixture signal $x[n]$ is a combination of respiratory and wheezing sounds, the magnitude spectrogram matrix of input mixture signal X is estimated as $\approx \hat{X} = \hat{X}_R + \hat{X}_W$. From any factorisation approaches the target sources are estimated by updating the decomposition models through update rules. Similarly in ST-NMPCF approach our target is to separate respiratory and wheezing sounds from input mixture signal $x[n]$. Therefore $\hat{X}_R$ and $\hat{X}_W$ estimated spectrograms of respiratory and wheezing sounds obtained through co-factorisation process and update rules which is discussed briefly in section 4.3.3

Segmentation of X into L-column blocks:

The main aim of our proposed approach is to collect both the spectral and temporal repetitive components of target source to improve the segregation performance. The key assumptions considered in our approach to proceed with the separation process are as follows:

1. It is concluded in [30] that respiratory sounds are usually characterized by their similar spectral patterns that represent a wideband noise spectrum which show smoothness in time and frequency domain. This particular characteristic of respiratory sounds is useful in segregating the commonly found spectro-temporal respiratory sounds by introducing training signal Y in the co-factorisation process.
2. Respiratory sounds are considered to have repetitive respiratory events in human breathing with very small variations [53]. This particular feature of respiratory sounds is helpful in modelling the common spectral components that are found throughout all the segmented frames. To take the advantage of this particular property of respiratory sounds, the magnitude spectrogram matrix of input mixture signal X is segmented in to L - column blocks $(X^{(1)}, X^{(2)}, X^{(3)}, \dots, X^{(L)})$ where each segment have similar spectral patterns of respiratory sounds in all of them. To achieve the process of segmentation we used AMIE_SEG [**95**] that automatically allows in segmenting the mixture spectrogram X into inspiratory and expiratory stages.

Decomposition of factor matrices

Every factorisation approach have their own spectrogram decomposition models which define about how the factorisation process is carried out

in that particular approach to reach their goals. Similarly in our proposed approach we too have our own factorisation models in which the magnitude spectrogram matrix X is decomposed in to frequency and time characteristic matrices based on common respiratory basis and harmonic wheezing basis vectors which is briefly discussed in section 4.3.2.2.

Weighting the factor matrices with ST-NMPCF decomposition and iterative update rules

The decomposed factor matrices obtained from factorisation model as explained above(discussed in section 4.3.2.2) are allowed to iterate for several iterations to make the decomposed matrices collect their respective sounds i.e. to end up respiratory basis vectors collecting all the spectral and temporal components of respiratory sounds and the harmonic wheezing basis vectors to collect all the individual components other than respiratory sounds i.e. wheezing sounds. The update rules are derived from the objective function of the proposed system (ST-NMPCF). The concept of objective function and the derivation of update rules are explained in section 4.3.3.

Reconstruction and concatenation of estimated matrices

When all the sound components are collected in their respective segmented matrices they are reconstructed by the process of Wiener filtering and concatenated together to form two separate signals $\hat{X}_R$ and $\hat{X}_W$.

These estimated spectrogram, matrices $\hat{X}_R$ and $\hat{X}_W$ are converted from frequency-domain to time-domain by Inverse Short-Time Fourier Transform to have the audio signals in time domain. Finally we result in two separated sounds $\hat{x}_R[n]$ (represents n-th sample of estimated respiratory sound) and $\hat{x}_W[n]$ (represents n-th sample of estimated wheezing sound). The process of reconstruction of the signals is explained briefly in section 4.3.4.

4.3.2 Factorisation Models

4.3.2.1 The concept of Factorisation models

Factorization models refer to the decomposition of a matrix into two matrices which on multiplying gives the approximate of the original matrix. For example, if we consider the magnitude spectrum of

respiratory training signal Y, then factorization model approximates the training signal by the product of two matrices A_R and S_R , which is represented by following equation:

$$Y \approx \hat{Y} = A_R S_R \qquad (4.3.2.1)$$

Where $\hat{Y}$ represents the estimated spectrogram of training signal. We can't recover the original spectrogram through factorization process, therefore whatever we obtain is a approximated version of the original one i.e. why it is represented with a variable approximate $\hat{Y}$. In general Matrix multiplication is implemented by computing the column vectors of Y as linear combinations of the column vectors in $A_R$using corresponding coefficients provided by columns of S_R. Where A_R and S_R represents the frequency and time characteristics of Y. As Y is a matirx of rows and columns, where each column of Y is approximated as follows:

$$\hat{Y}_i = A_R S_{Ri} \qquad (4.3.2.2)$$

$\hat{Y}_i$ from equation (4.3.2.2) represents the approximated i^{th} column vector of the training signal spectrogram Y, which is obtained by multiplying A_R with the corresponding column-vector S_{Ri}. From this explanation we want to give an idea that the magnitude spectrogram of any matrix can be represented by product of two matrices where rows of one matrix contains the spectral components of the respective signal and the columns of other represents its respective time-domain activations. From equation (4.3.2.2) A_R contains the spectral components of respiratory signal and the time-domain activations of each row are contained in each column vector of the S_R matrix.

The concept of representing a single-matrix in a product of two different matrices is to reduce the time and space complexity of the computation process. Moreover, when matrices are multiplied, the dimensions of the factor matrices will be significantly smaller than those of the product matrix and this property of matrices forms the basis of the matrix factorization approaches like NMF, NMPCF and many others. The process of factorization significantly reduces dimensions compared to the original matrix. Similarly, the ST-NMPCF approach also follows a different factorization model which plays a key role in explicitly segregating the respiratory sounds and wheezing sounds from the input mixture signal $x[n]$.

4.3.2.2 Factorisation models for ST-NMPCF approach

As the main aim of our proposed approach is to extract both the spectral and temporal components of the respiratory sounds we consider respiratory training signal along with the input mixture signal to remove the respiratory sounds from input mixture signals. As described in section 4.3.1 if the respiratory training signal is represented as $y[n]$ in its time-domain then Y represents its magnitude spectrum of the signal in its frequency domain. Similarly if the respiratory mixture signal is represented by $x[n]$ then in its magnitude spectrum is represented as X in its frequency domain. The magnitude spectrogram matrices of the respiratory training signal Y and the segmented column-blocks of the input mixture signal X are decomposed as shown in equation (4.3.2.3) and equation (4.3.2.4). These matrices by using multiplicative update rules (explained in section 4.3.3) and partial co-factorisation process learn the basis vectors of respiratory and wheezing components which on reconstruction provide us the separated or estimated respiratory and wheezing sounds spectrograms $\hat{X}_R$ and $\hat{X}_W$.

$$Y \approx \hat{Y} = \hat{X}^{(1)} = A_R S_R \qquad (4.3.2.3)$$

$$X^{(l)} \approx \hat{X}^{(l)} = \hat{X}_R^{(l)} + \hat{X}_W^{(l)} = A_R S_R^{(l)} + A_W^{(l)} S_W^{(l)} \qquad (4.3.2.4)$$

Where l in equation (4.3.2.4) is of range $(2 < l < L)$. In equation (4.3.2.3) the estimated magnitude spectrogram of respiratory training signal $\hat{Y}$ is represented as $\hat{X}^{(1)}$ for simplicity in writing the algorithm. To take the advantage of temporal source separation process the mixture signal X has to be segmented to L-column blocks which contain repetitive respiratory events as explained in section 4.3.1, where the magnitude spectrogram of l^{th} segmented block is represented by $X^{(l)}(X^{(1)}, X^{(2)}, X^{(3)}, \ldots, X^{(L)})$. From equation 4.3.2.3 every column vector in the matrix Y is approximated by a weighted linear combination of the respiratory basis vectors in the columns of A_R, the weights for basis vectors(the time-domain activations) appear in the corresponding column of the matrix S_R. As magnitude spectrogram matrix of training signal Y contains solo-respiratory sounds, A_R is estimates or collects all the spectral components solo-respiratory sounds which are commonly found in all individual subjects. The spectrogram of input mixture signal X is composed of both respiratory and wheezing sounds but we don't have the basis vectors of respiratory sounds and wheezing sounds before hand, we have to estimate them simultaneously through partial

co-factorisation process. To estimate the basis vectors of respiratory and wheezing sounds the spectrogram of input mixture signal X is approximated as a linear combination of sum of respiratory and wheezing basis vectors . A_R is assigned for collecting all basis vectors of repetitively found respiratory events in all the segmented blocks. The matrix A_R used in decomposition models equation (4.3.2.3) and in equation (4.3.2.4) are same. By assigning same A_R to both the models A_R collects the spectral characteristics of breath sounds that is obtained from the training signal Y as well as it also collects the spectral components of temporally repetitive respiratory sounds present in all the segmented blocks of the respiratory mixture signal X. This process helps in removing all the respiratory components from the mixture signal X and improve the audibility of the wheezing sounds. The matrix $S_R^{(l)}$ contains the time domain activations of respiratory sounds of l^{th} segmented block. Similarly the estimated wheezing spectrogram of l^{th} segmented block $\hat{X}_W^{(l)}$ is obtained by learning the basis vectors of the wheezing sound through $A_W^{(l)}$ matrix and its respective time domain activation are obtained from $S_W^{(l)}$ matrix. All matrices A_R, $A_W^{(l)}$, $S_R^{(l)}$ and $S_W^{(l)}$ are constrained to be non-negative.

To learn the basis matrices of respiratory and wheezing sounds A_R and $A_W^{(l)}$, the matrices have to be allocated with correct number of respiratory and wheezing components to separate the target sources from the mixture signal X. Therefore matrix A_R is dimensioned to a size of $F \times K_R$, where F represents the total number of frequency components of the input mixture signal X and K_R represents the number of respiratory components that have to be decided by the user to separate the respiratory sounds from the mixture signal X. Similarly $A_W^{(l)}$ is dimensioned to a size of $F \times K_W$, where K_W represents the number of wheezing components that have to be decided by the user to separate the wheezing components from the X. Contrastingly $S_R^{(l)}$ $(1 < l < L)$ is dimensioned to a size of $K_R \times T$ where T represents the number of time domain activation of respiratory sounds in each of the l^{th} segmented block. Whereas $S_W^{(l)}$ $(2 < l < L)$ is dimensioned to a size of $K_W \times T$ where T represents the number of time domain activation of wheezing sounds in each of the l^{th} segmented block. Wheezing components are found only in input mixture signal whereas respiratory components are

found commonly in both training signal and input mixture signal, that is why the range of l is different for $S_R^{(l)}$ and $S_W^{(l)}$ Since $l = 1$ represents the respiratory training signal ($Y \approx \hat{X}^{(1)}$). ***Figure 4.3.2*** depicts the pictorial representation of the decomposition models of the respiratory training signal Y and input mixture signal X for the proposed ST-NMPCF approach.

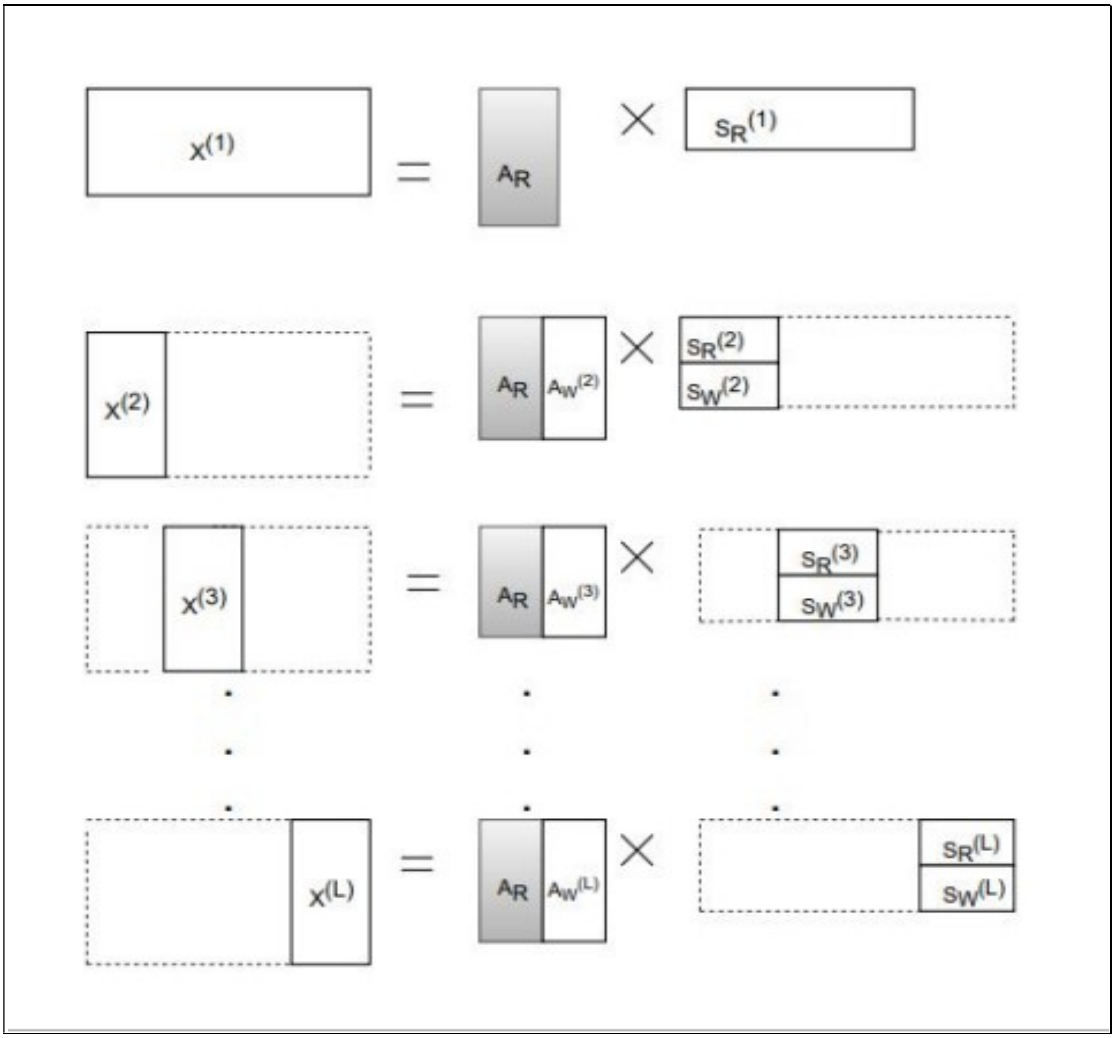


Figure 4.3.2 Pictorial representation of decomposition models of respiratory training signal $Y = X^{(1)}$ and input mixture signal X for proposed ST-NMPCF approach

Finally to summarise, from equation (4.3.2.3) the basis matrix A_R estimates the spectral components of all the respiratory sounds by learning the spectral characteristics of respiratory sounds from the respiratory training signal. By allowing the same matrix A_R to get partially co-factored in equation (4.3.2.4) the basis matrix takes the

advantage and learns temporal components of repetitive respiratory sounds found in segmented column blocks $(X^{(2)}, X^{(3)}, \dots, X^{(L)})$ of input mixture signal's magnitude matrix X. As a result the matrix AR ends up in collecting all the spectral and temporal components of the respiratory sounds. However $A_W^{(l)}$ is responsible for learning the individual wheezing components that are present in the L-segmented blocks. All the matrices are considered to be non-negative. This is how the factorisation models designed assist us to learn and collect the respective sources that we want to separate from mixture.

4.3.3 Objective Function and Update Rules

In order to estimate the matrices of decomposed models in equation 4.3.2.3 and 4.3.2.4 first we have to solve the minimisation problem. This is solved by minimising the cost function between original and estimated matrices. There are different cost functions like least squares method, α-divergence and β-divergence. We have discussed some of these while explaining NMF in state of art methods in section 2.5.1. In our proposed methodology the objective function we used is generalised using the concept of β-divergence.Therefore for given L non-negative input matrices $X^{(l)}$ $(1 < l < L)$. ST-NMPCF seeks to minimise the objective function by following equation (4.3.3.1).

$$\Gamma_{ST-NMPCF} = \sum_{l=1}^{L} \lambda_l \left\| X^{(l)} - A_R S_R^{(l)} - A_W^{(l)} S_W^{(l)} \right\|_F^2 + \gamma \left\{ \sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2 \right\}$$
$$= \sum_{l=1}^{L} \lambda_l D_F\left(X^{(l)} | A_R S_R^{(l)} + A_W^{(l)} S_W^{(l)}\right) + \gamma D_F(A_R|0) + \gamma \sum_{l=1}^{L} D_F\left(A_W^{(l)}|0\right) \quad (4.3.3.1)$$

$\sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2$ is the regularization term defined as $\sum_{l=1}^{L} \left\| A^{(l)} \right\|_F^2 = L\|A_R\|_F^2 + \sum_{l=1}^{L} \left\| A_W^{(l)} \right\|_F^2$ and D_F is the representation for frobenius form. The β-divergence is defined as:

$$d_\beta(x|y) = \frac{1}{\beta(\beta-1)} \left(x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}\right), \beta \epsilon \mathbb{R}\{0,1\} \quad (4.3.3.2)$$

The special cases of β-divergence are explained in detail in equation 2.5.6 and equation 2.5.7. By applying the β-divergence on the objective function of ST-NMPCF approach we obtain:

$$\Gamma_{ST-NM}^{\beta} = \sum_{l=1}^{L} \lambda_l D_\beta\left(X^{(l)} | A_R S_R^{(l)} + A_W^{(l)} S_W^{(l)}\right) + \gamma L D_F(A_R|0) + \gamma \sum_{l=1}^{L} D_F\left(A_W^{(l)}|0\right) \quad (4.3.3.3)$$

Definition of the terms used in objective function:

λ_l :- This term contributes to control the reconstruction error of each input matrix among column-blocks and training data with solo-respiratory sounds.

D_β :- Represents the β-divergence function

D_F :- Represents the Frobenius form i.e $D_F(A|B)$ can be represented in its Frobenius form as $A - B$.

L :- Represents the number of segmented column-blocks of input mixture signal $x[n]$.

γ :- Regularization parameter to avoid basis vectors from converging them to small values i.e. to maintain non-negativity of the matrices.

A_R :- Matrix which collects the spectro-temporal components of R common basis vectors(respiratory sounds)

$S_R^{(l)}$:- l^{th} column matrix of input mixture signal X which estimates the weights or time-domain activations of respiratory sounds of A_R, where l is of range $1 \leq l \leq L$.

$A_W^{(l)}$:- l^{th} column matrix of input mixture signal X which estimates the individual basis vectors (wheezing sounds), where l is of range $2 \leq l \leq L$.

$S_W^{(l)}$:- l^{th} column matrix of input mixture signal X which estimates the weights or time-domain activations of respiratory sounds of $A_W^{(l)}$, where l is of range $2 \leq l \leq L$.

After obtaining the objective function the next step is to derive multiplicative update rules for each matrix that has to be estimated. Multiplicative update rules will be based on gradient descent. In traditional gradient descent, the local minima of the cost function is found by moving in the direction of its gradient, i.e., its direction of steepest descent. At each iteration, we move along the independent variables of the cost function so that, after a certain learning rate, the cost function gets decreased maximally. The gradient for each matrix is obtained by decomposing the gradient of above objective function with

respect to particular matrix. For example if we want to get gradient for matrix A_R then $\frac{\partial \Gamma^{\beta}}{\partial A_R}$ is represented as follows:

$$\frac{\partial \Gamma^{\beta}}{\partial A_R} = \left[\frac{\partial \Gamma^{\beta}}{\partial A_R}\right]^{+} - \left[\frac{\partial \Gamma^{\beta}}{\partial A_R}\right]^{-} \qquad (4.3.3.4)$$

Then the multiplicative update rules are derived for particular matrix A_R by:

$$A_R \longleftarrow A_R \odot \frac{\left[\frac{\partial \Gamma^{\beta}}{\partial A_R}\right]^{-}}{\left[\frac{\partial \Gamma^{\beta}}{\partial A_R}\right]^{+}} \qquad (4.3.3.5)$$

Note : We can calculate the derivative of $D_{\beta}(x|y)$ with respect to y as follows:

$$\frac{\partial D_{\beta}(x|y)}{\partial y} = y^{\beta-2}(y-x)$$

And this leads to the following gradients of the objective functions

$$\frac{\partial \Gamma^{\beta}_{ST-NMPCF}}{\partial A_R} = \sum_{l=1}^{L} \lambda_l \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)}\right)^{.(\beta-2)} \odot \left(A_R S_R^{(l)} + A_W S_W^{(l)} - X(l) \times SR(l)T + 2\gamma LAR \right. \right. \qquad (4.3.3.6)$$

$$\frac{\partial \Gamma^{\beta}_{ST-NMPCF}}{\partial A_W^{(l)}} = \lambda_l \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)}\right)^{.(\beta-2)} \odot \left(A_R S_R^{(l)} + A_W S_W^{(l)} - X(l) \times SW(l)T + 2\gamma AW(l) \right. \right. \qquad (4.3.3.7)$$

$$\frac{\partial \Gamma^{\beta}_{ST-NMPCF}}{\partial S_R^{(l)}} = \lambda_l A_R^T \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)}\right)^{.(\beta-2)} \odot \left(A_R S_R^{(l)} + A_W S_W^{(l)} - X(l) \right. \right. \qquad (4.3.3.8)$$

$$\frac{\partial \Gamma^{\beta}_{ST-NMPCF}}{\partial S_W^{(l)}} = \lambda_l \left(A_W^{(l)}\right)^T \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)}\right)^{.(\beta-2)} \odot \left(A_R S_R^{(l)} + A_W S_W^{(l)} - X(l) \right. \right. \qquad (4.3.3.9)$$

Henceforth the multiplicative update rules for all the matrices are derived by taking positive terms of the partial derivative of the objective function as denominator of multiplication factor, while taking negative terms on numerator as described in equation (4.3.3.5) By using the

equations (4.3.3.6 to 4.3.3.9) the update rules for each matrix are derived as follows:

$$A_R \leftarrow A_R \odot \frac{\sum_{l=1}^{L} \lambda_l \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-2)} \odot \left(X^{(l)} \right) \right\} \times \left(S_R^{(l)} \right)^T}{\sum_{l=1}^{L} \lambda_l \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-1)} \right\} \times \left(S_R^{(l)} \right)^T + 2\gamma L A_R} \quad (4.3.3.10)$$

$$A_W^{(l)} \leftarrow A_W^{(l)} \odot \frac{\lambda_l \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-2)} \odot \left(X^{(l)} \right) \right\} \times \left(S_W^{(l)} \right)^T}{\lambda_l \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-1)} \right\} \times \left(S_R^{(l)} \right)^T + 2\gamma A_W^{(l)}} \quad (4.3.3.11)$$

$$S_R^{(l)} \leftarrow S_R^{(l)} \odot \frac{A_R^T \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-2)} \odot \left(X^{(l)} \right) \right\}}{A_R^T \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-1)} \right\}} \quad (4.3.3.12)$$

$$S_W^{(l)} \leftarrow S_W^{(l)} \odot \frac{\left(A_W^{(l)} \right)^T \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-2)} \odot \left(X^{(l)} \right) \right\}}{\left(A_W^{(l)} \right)^T \left\{ \left(A_R S_R^{(l)} + A_W S_W^{(l)} \right)^{.(\beta-1)} \right\}} \quad (4.3.3.13)$$

The multiplicative update rules have an important property of not allowing any of the matrices to lose their non-negativity constraint during the iteration process.

4.3.4 Separation Procedure

Table-1 explains the procedural steps followed to implement the ST-NMPCF approach. In the factorization model of S-NMPCF, there are only two matrices Y (prior-side information of solo-respiratory sounds), X (input respiratory mixture signal) (Fig 4.1.1). An important thing to note here is that S-NMPCF approach does not include any segmentation process. Whereas T-NMPCF approach is a unsupervised approach and does not use any training signal, but the input spectrogram of input mixture signal X is segmented into L consecutive-column blocks without considering. As a result, we get L number of consecutive matrices. This model takes the advantage of the occurrence of repeatable components in the signal and tries to separate them from the mixture signal. Finally by taking the advantages of both the models (S-NMPCF and T-NMPCF), our proposed ST-NMPCF for source separation was programmed to attain the best results. In our proposed model, respiratory training signal $y[n]$ formed by concatenating solo-respiratory sounds from database D3. while the rest (L-1) matrices are the segmented

column-blocks of the input-mixture signal. In S-NMPCF and proposed ST-NMPCF source separation approaches $A_W^{(1)}$and $S_W^{(1)}$ matrices are initialized with empty matrices. These are initialized with empty matrices because the training database which consists of solo-respiratory sounds is responsible to make $A_R$learn the spectral information of respiratory sounds to that in input mixture signals but it won't hold any information regarding wheezing components, therefore they are set to zero.

It is hard to estimate the phase information of each segmented column matrix during the update process. Therefore, in order to overcome this difficulty, the phase information of the L segmented column blocks of the input-mixture signal X is set aside to use them during reconstruction process. $\Phi(l)$ $(1 \leq l \leq L)$ represents the phase values of each column-matrix. These values are reused during the reconstruction processs of each estimated column block. In ***Table-4.3.1***, the 3rd step explains about the reconstruction process of each column-block, where the pre-stored phase information is used to reconstruct the original signal. There is an alternative procedure that can be followed to attain even more natural-sounding reconstructions i.e. by using Wiener filtering. At last, the reconstructed column blocks are concatenated end to end to get separated spectrograms of wheezes and respiratory sources. Now the spectrograms are resynthesized into their time-domain representation by applying inverse STFT.

Table 4.3.1 Algorithm for ST-NMPCF Source Separation Approach

---

1. Required signals and parameters:
   $y[n]$:- Respiratory training signal
   $x[n]$:- Input respiratory mixture signal

   $K_R, K_W, \lambda_l, \gamma$
2. Preparation of matrices
   a. Calculate the magnitude spectrogram Y of training signal $y[n]$ by using equation 4.3.1.1 and consider it as $Y = X^{(1)}$.
   b. Calculate the magnitude spectrogram X of input respiratory mixture signal $x[n]$ by using equation (4.3.1.1) and segment it into column blocks of inspiratory and expiratory phases by AMIE_SEG [95].

c. Save the phase information of respiratory training signal Y and L segmented column-blocks of mixture signal X, $(2 \leq l \leq L)$.
d. Initialize the factor matrices A_R, $A_W^{(l)}$, and $S_W^{(l)}$ $(2 \leq l \leq L)$ with random positive values.
e. Initialize the factor matrices $S_R^{(l)}$ $(1 \leq l \leq L)$ with random positive values.
f. Initialize the factor matrices $A_W^{(1)}$, and $S_W^{(1)}$ with empty matrices.

3. Update each factor matrix using update rules derived in equations (4.3.3.10, 4.3.3.11, 4.3.3.12, 4.3.3.13) for the predefined number of iterations.
4. Reconstruct the estimated spectrograms.
   a. Reconstruct the spectrogram of the wheezing sources and respiratory sources in each column block by following one of the two approaches

   $\hat{X}_R^{(l)} = \left(A_R S_R^{(l)}\right) \odot \Phi(l) \quad (2 \leq l \leq L)$

   $\hat{X}_W^{(l)} = \left(A_W^{(l)} S_W^{(l)}\right) \odot \Phi(l) \quad (2 \leq l \leq L)$

   Or alternatively, Wiener filtering can be used like:

   $$\hat{X}_R^{(l)} = X^{(l)} \odot \frac{A_R S_R^{(l)}}{A_R S_R^{(l)} + A_W^{(l)} S_W^{(l)}} \quad (2 \leq l \leq L)$$

   $$\hat{X}_W^{(l)} = X^{(l)} \odot \frac{A_W^{(l)} S_W^{(l)}}{A_R S_R^{(l)} + A_W^{(l)} S_W^{(l)}} \quad (2 \leq l \leq L)$$

5. Concatenate the reconstructed column-blocks:

   $\hat{X}_R = \hat{X}_R^{(2)} + \hat{X}_R^{(3)} + \cdots \hat{X}_R^{(L)}$

   $\hat{X}_W = \hat{X}_W^{(2)} + \hat{X}_W^{(3)} + \cdots \hat{X}_W^{(L)}$
6. Inverse-transform the matrices $\hat{X}_R$and $\hat{X}_W$ into time-domain to get separated estimated respiratory audio signal $\hat{x}_r[n]$ and separated wheeze signal $\hat{x}_w[n]$.

---

5 EXPERIMENTAL RESULTS

5.1 Database

Respiratory diseases are found to be an increasing health threat nowadays and many researches were undergoing which reflect on rapid and significant progress in pulmonary science. However there were no publicly available internet pulmonary repositories with which the newly developed algorithms can be evaluated and compared. Moreover it is also impossible to find pure-respiratory sounds on the internet. It is difficult to have a large database with all respiratory and wheezing sounds. Therefore it is a challenging task to create one to evaluate the proposed methodology. This has been complained as a major drawback by authors who are researching lung sound analysis and separation of abnormal sounds from the given mixture signal in biomedical signal processing. Every researcher had been creating their own dataset to evaluate their experimental results which were a challenging task. Therefore with all this inconvenience of not having a proper pulmonary sound database we created three datasets D1, D2 and D3 with a total of 71 audio files. D1 and D2 combined contain 64 respiratory audio samples and are considered to be one complete database D which contain several Respiratory mixture signals. We separated the dataset D and assigned each dataset to evaluate different tasks in the proposed ST-NMPCF approach. D1 is composed of 48 which is ¾ of the database D whereas ⅓ of database D are assigned to D2 which is composed of 16 audio file recordings and D3 is composed of 7 audio file recordings. The dataset D1 is used for the optimisation process. The details of the Optimisation process is explained clearly in section 5.3. Dataset D2 is used for the testing process and details about the process are explained in section 5.4. The dataset D3 is used for creating the training data signal. The dataset D3 consists of audio recordings of solo-respiratory signals i.e. without any abnormal sounds in it. There are seven audio recording files which are concatenated end-to-end to form a training data signal or prior side information which relates to y[n] in our approach. The details about these files and their characteristics are explained clearly in ***Table-5.1.1.***

All these datasets are created by collecting the audio files from the mostly used and available online pulmonary repositories [55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, and 67]. The important point to note about these datasets is, one dataset is not a part of another dataset. To be more specific any file that is used in D1 is not used in the other two datasets and vice-versa with other two databases. All the audio files we collected from the internet are recorded with the help of a stethoscope or microphone. These files include sounds from both healthy and unhealthy individuals which are recorded from various lung sound occurrence points i.e. at anterior chest, the mid axillary region, and the posterior chest of the individuals. These files include sounds from unhealthy individuals who are suffering with asthma; COPD and bronchitis in whose breath the occurrence of wheezes are found to be common.

Dataset	**D1**	**D2**	**D3**
No. of Files	48	16	7
Purpose	Optimisation	Testing	Training
Total Time of recordings	480 sec	146 sec	90 sec
Min-Max length of audio files	2.3 sec - 8.2 sec	3 sec to 14 sec	8 sec - 20 sec
No. of RS events	140	65	30
No. of WS	96 to 100	17	---

Table 5.1.1 Characteristics of each Database D1, D2 and D3

In table 5.1.1 RS represents respiratory sound and WS represent Wheezing sound. All the signals of the created databases D1, D2, D3 are featured as single channel type with sampling frequency of 2048 Hz with a resolution of 16-bit. All the three databases are created by us by collecting several respiratory sounds from the internet.

Steps followed to create the dataset:

1. Collected all files of respiratory sounds available on the internet keeping in mind that all these sounds are recorded through a stethoscope or microphone for better performance results.
2. According to the characteristics of wheezing sounds, these are sounds that are superimposed on normal respiratory breaths, and we get these sounds along with the respiratory breath signal. As it is difficult to get solo wheezing sounds from the internet we decided to separate the wheezes manually from the collected respiratory mixtures. The manual separation process is explained in the later part of the document.
3. After getting the wheezing sounds w[n] and respiratory sounds r[n] we mixed the respiratory file with manually separated wheezing sound files to get the mixture signal.
4. From the collected respiratory sounds we left few for creating the respiratory training signal that we named it as D3.
5. Having the respiratory mixtures from step 3 we randomly selected some file to create the dataset for optimization process i.e. D1 and we left the other mixtures for testing purpose i.e. named as D2.
6. A very important point to be noted here is that no dataset includes audio recording files from another dataset. All the files in each dataset are different from each other.

Manual wheezing sound separation process.

The process of manually separating the wheezing sounds from mixture signals to create a database of our own is a very tedious and time-consuming task to implement. We used MATLAB as a tool to implement this task. With MATLAB we can have a clear visualisation of the spectrogram representation of the signal. The spectrogram representation of a signal gives the frequency bins of active wheezing and respiratory sounds. With this data we tried to make frequency bins of active respiratory sounds to zero leaving the wheezing components behind. The wheezing components are considered along with their phase information. These are also tested with listening activity to check for the sound quality. Finally we ended up with wheezing sound files with all its respiratory sounds inactive.

While creating the dataset D1 we made three different sets (D1 5dB, D1 0dB, and Dl -5DB) with the same respiratory (r[n]) and wheezing (w[n]) files. The difference between these three data sets is these sounds have

been mixed considering different Signal to Noise Ratio (SNR). To mention them more clearly

- D1 5dB signals are created by considering the power of wheezing sound signal w[n] 5dB greater than respiratory sound signal r[n]. This allows us to hear wheezing sounds louder than the respiratory sound.
- D1 0dB signals are created by considering the power of wheezing sound signal w[n] and respiratory sound signal r[n] as the same value. This allows us to hear wheezing sounds and the respiratory sounds with the same loudness.
- D1 -5dB signals are created by considering the power of wheezing sound signal w[n] 5dB less than the respiratory sound signal r[n]. This allows us to hear respiratory sounds louder than the wheezing sounds.

We created these datasets to check for performance efficiency of the signals. Because we don't have any idea of how the respiratory sounds will be when they are getting recorded. Therefore to check the performance with all sorts of possibilities we created these datasets. But in all our experimental evaluation we used only D1 0dB dataset to evaluate the optimisation results in order to obtain unique optimization parameters taking into account the intermediate SNR i.e. 0dB.

5.2 Metrics

Metrics are measures that are commonly used for assessing, comparing and evaluating the performance of any methodology. In our proposed methodology we used a tool to compute the metrics whose values allow us to compare the performance of the different methodologies with the proposed methodology. BSS Eval toolbox [68][69] is one of the toolboxes that is used in MATLAB that is designed by authors to measure the performance of audio source separation algorithms within an evaluation framework where the original source signals are available as ground truth. The measures are based on the decomposition of each estimated source signal into a number of contributions corresponding to the target source, interference from unwanted sources, and artifacts such as noise. The BSS Eval tool allows us in evaluation of three metrics.

1. SDR (Source to Distortion Ratio): SDR gives the overall performance measure of the separation methodology used.
2. SIR (Source to Interference Ratio): SIR gives the information on the interferences of other sounds in the signal. For example the interference of wheezing sounds in respiratory sounds and vice-versa.
3. SAR (Source to Artifacts Ratio): SAR gives information on artifacts if there are any in the separated signals.

Separation performance measures are computed for each estimated source by comparing it to a given true source. If $s_t[n]$ is the true source and $s_e[n]$ is the estimated source of the signal, the computation of these matrices is done in two steps. First step is to express the estimated source in terms of its errors i.e. shown in equation below:

$$s_e[n] = s_t[n] + e_{interf}[n] + e_{spatial}[n] + e_{artifacts}[n] \quad (5.2.1)$$

$$s_e[n] - s_t[n] = e_{interf}[n] + e_{spatial}[n] + e_{artifacts}[n] \quad (5.2.2)$$

$e_{interf}[n]$ is an error term that accounts for the interferences of the unwanted sources, $e_{spatial}[n]$ is an error term that accounts for spatial distortion, and $e_{artifacts}[n]$ is an error term that are related to artifacts of the separation algorithm or simply related to deformations introduced by the separation algorithm that are not allowed. There are many ways through which we can compute such a decomposition depending on which transformations are allowed, and the toolbox includes the function to evaluate those. More details about the BSS evaluation tool is described briefly in [68] [69] [70].

As discussed the computation of metrics follows two steps, we already discussed the first step i.e. to decomposes$s_e[n]$, whereas the second step is to define numerical performance criteria of each of the three metrics by computing energy ratios expressed in decibels (dB).

Source to Distortion Ratio is defined by :

$$SDR = 10\log_{10}\frac{\|s_e[n]\|^2}{\left\|e^e_{interf}[n]+e^e_{spatial}[n]+e^e_{artifacts}[n]\right\|^2} \quad (5.2.3)$$

Source to Interference Ratio is defined by :

$$SIR = 10\log_{10}\frac{\|s_e[n]\|^2}{\left\|e^e_{interf}[n]\right\|^2} \quad (5.2.4)$$

Source to Artifacts Ratio is defined by :

$$SAR = 10\log_{10}\frac{\left\| s_e[n]+e^e_{interf}[n]+e^e_{spatial}[n] \right\|^2}{\left\| e^e_{artifacts}[n] \right\|^2} \quad (5.2.5)$$

Source to Noise ratio is defined by:

$$SNR = 10\log_{10}\frac{\left\| s_e[n]+e^e_{artifact\ }[n] \right\|^2}{\left\| e^e_{spatial}[n] \right\|^2} \quad (5.2.6)$$

$s_e[n]$ represents the signal that we want to evaluate the metrics for. In our proposed methodology we have to evaluate the metrics for estimated wheezing sound $\hat{x}_w[n]$ and estimated respiratory sound $\hat{x}_r[n]$ as indicated in flow chart figure 4.3.1.1. Therefore $s_e[n]$ can be $\hat{x}_r[n]$ or $\hat{x}_w[n]$. The indication of terms will be different while analysing different target sounds.

For wheezing signal:

$s_e[n] = \hat{x}_w[n]$ (the separated estimated wheezing sound extracted by using the ST-NMPCF separation algorithm.).

$s_t[n] = x_w[n]$ (the original wheezing signals that are obtained from created database while creating the mixture signals).

The other terms are indicated by

$\left(e^e_{interf}[n], e^e_{spatial}[n], e^e_{artifacts}[n]\right) = \left(e^w_{interf}[n], e^w_{spatial}[n], e^w_{artifacts}[n]\right)$

For respiratory signal:

$s_e[n] = \hat{x}_r[n]$ (the separated estimated wheezing sound extracted by using the ST-NMPCF separation algorithm.).

$s_t[n] = x_r[n]$ (the original wheezing signals that are obtained from created database while creating the mixture signals).

The other terms are indicated by

$\left(e^e_{interf}[n],\ e^e_{spatial}[n],\ e^e_{artifacts}[n]\right) = \left(e^r_{interf}[n],\ e^r_{spatial}[n],\ e^r_{artifacts}[n]\right)$

By using above mentioned information and computational equations we ran the BSS evaluation tool in MATLAB and we obtained two different sets of values for each metric SDR, SIR, SAR, SNR. They are as follows:

1. SDR_r, SIR_r, SAR_r, SNR_r that refer to values of respiratory sounds
2. SDR_w, SIR_w, SAR_w, SNR_w that refer to values of respiratory sounds.
3. Higher values of SDR, SIR, SAR, SNR indicates higher audio quality of estimated sounds

5.3 Experimental Setup

In this section we are going to discuss the assumed parameters and experimental environments we considered to evaluate the separation performance and computational cost of the proposed methodology and its comparison methods.

About audio recordings:

The audio files we considered for training signal and input mixture signal are all from created database that has been explained in section 5.1

Channel type:

All the audio signals used for the experimental purpose are of single or mono channel type.

Sampling rate and bits per sample:

All the input signals used for experimenting proposed methodology are sampled at a rate of 2048 Hz, and are encoded in 16 bits per sample just as in ordinary PCM (Pulse Code Modulation).

Transform:

All signals with sampling rate of fs= 2048 Hz are transformed into their frequency domain representation by considering a sample length of N = 256, with an overlap of 25% of sample length which result in Hop samples of 64. The windowing technique used here is Hamming window and discrete Short time fourier transforms (nfft) of 2N points is considered. Number of samples that have no overlap are (N- Hop samples = 192).

Training signal:

As explained in section 5.1 a database was created to use it as a training signal to evaluate the performance of ST-NMPCF approach. The

training signal used is a concatenation of seven respiratory signals from D3 database which account to a time of 90 sec approximate. A point to note here is we don't consider any training data for T-NMPCF approach.

γ:

Gamma is the regularization parameter. It is used to control or prevent the basis vectors from converging them into too small values. But its value is set to 1 in all cases because of its less importance in comparison of the systems. This is because this value does not influence the separation results much. In general regularization parameters are added to optimize the separation performance among different approaches. In our case we compare the separation performance among the approaches S-NMPCF, T-NMPCF, and ST-NMPCF, but variation of gamma among these approaches hasn't shown any improvements in the separation results. Therefore the value is set to 1.

λ_l:

This term accounts for controlling the reconstruction error among each column-block of the input matrix and the respiratory training signal. To determine the value of parameter λ which decides the relative importance between input mixture signal and respiratory training data, we considered different values of λ. The values considered are detailed in optimization process.

Number of Iterations:

One of the main ambiguities of factorisation models in source separation tools is, the convergence of objective function does not guarantee best separation performance. It is also proved often that too many iterations result in over fitting, contrastingly not enough iteration provide bad or worse results. Therefore we basically iterate all NMPCF algorithms for 100 times to investigate the performance oscillations that are being caused by iterations.

Number of Respiratory basis vectors (K_R) and individual wheezing components (K_W):

Basis vectors represent the number of components of respective target in the mixture signal. The selection of optimal number of basis vectors is an important ambiguity found in matrix factorization methods used in

source separation algorithms. These values are needed to be approximated for proper separation of signals and its reconstruction. Therefore in our experiment we took various number of basis vectors for K_R and K_W and checked for different combinations to select the optimal number that result in maximum removal of respiratory sounds from the mixture signal.
The performance of the proposed methodology depends on how properly we estimate or initialize the K_R and K_W values. To check the performance we evaluated the algorithm for the times and the average of the values is considered as results.

5.4 Comparison Methods

In order to show that the proposed NMPCF approach gives better separation performance than other two NMPCF approaches (S-NMPCF and T-NMPCF). We evaluated the database D1 with all the three approaches to compare the separation efficiency. Based on the optimal parameters that define the particular algorithm for separation of target source from the mixture signal we did this comparison process. We checked the novelty of the proposed method of NMPCF with the other approaches. We took the following considerations.

- We considered the same database D1 to run all the NMPCF algorithms along with the proposed method.
- Only the proposed method ST-NMPCF and S-NMPCF approach consider training data $y[n]$ for the evaluation process. The training data $y[n]$ is formed by concatenating the signals from database D3.
- The T-NMPCF approach does not consider any training data to compute with the evaluation process.
- The T-NMPCF and ST-NMPCF approaches are similar except the fact that we don't consider the training signal in T-NMPCF approaches.

5.5 Optimization

The main goals of the optimisation process is to minimise the cost function and maximise the separation performance by evaluating the optimal parameters for a given mixture signal. As seen in section 4.3.3

while explaining the objective function and update rules the proposed method depends on different parameters to improve the separation efficiency and more specifically to improve the audibility of estimated wheeze sound $\hat{x}_w[n]$. It is also explained that it is very important to estimate the number of respiratory and wheezing components K_R and K_W in the mixture signal for proper sharing of basis vectors between common and individual components. So, in order to obtain optimal parameters which provide the best performance from the signals of database D1, the algorithm has been tested on with different parameters to attain the optimal combination of the parameters.

Our proposed methodology depends on the parameters K_R, K_W, λ_l and γ. The considered values for each of the parameters to obtain the optimal combination are as follows: K_R**:**[8,16,32,64,128,256,512]**,** K_W :[8, 16, 32, 64, 128,256, 512]**,** λ_l: [0.001, 0.01, 0.1, 1] and $\gamma = 1$. Table 5.5.1 gives the clear overview of considered parameters set up for the optimization process.

Parameter	**Value**
K_R	[8, 16, 32, 64, 128, 256, 512]
K_W	[8, 16, 32, 64, 128, 256, 512]
λ_l	for $l = 1$ λ_l = [0.001, 0.01, 0.1, 1] for $l > 1$ λ_l = 1
γ	1
Number of iterations	100

Table 5.5.1 Parameter setup for evaluating the proposed source separation approach by varying different parameters

The whole optimization process is done for all the mixture signals of database D1 to obtain the best estimated wheezing signal $\hat{x}_w[n]$ along with the information of the metric values like, SDR_W, SIR_W, SAR_W and SNR_W. All the mixture signals of the database are evaluated with all the possible combinations of the specified parameter values described in ***Table 5.5.1***. In order to optimise the results we segmented the mixture

signal by using automatic segmentation method called AMIE_SEG that segments the respiratory mixture signal spectrogram X as inspiratory and expiratory stages. Respiratory sounds differs from person to person, which results in difference in lengths of inspiratory and expiratory stages. Therefore manually entering different length of time segment while examining each and every test mixture signal is a time-consuming task. So, we used this automatic segmentation process to segment the mixture signal spectrogram into L number of segments to collect the temporally repetitive respiratory characteristics of respiratory events in the signal. ***Table 5.5.2*** gives information about the optimal parameter values that provided the best estimated wheezing signal $\hat{x}_w[n]$ with enhanced audio quality of wheezing sound.

Parameter	K_R	K_W	λ_1	γ	λ_l $(l > 1)$
Optimal value	128	8	0.001	1	1

Table 5.5.2 Optimal parameter values obtained for best wheezing sound separation through proposed approach achieved by evaluating Database D1

The optimal values are decided by evaluating the SDR_W values of estimated wheeze signal compared to original wheeze sound that was used to form a mixture signal. Source to Distortion ratio gives the overall separation quality of the proposed source separation approach (ST-NMPCF). We ran our algorithm by varying different values of parameters to check the stability of the proposed algorithm with the variations applied.

Figure 5.5.1 depicts the SDR_W results of best wheeze separated signal obtained by doing the optimisation process on all signals of Database D1. The SDR_W values obtained by different combinations of values taken for parameters K_R and K_W as mentioned in table 5.5.2. The values of $\lambda_1 = 0.001$ and $\lambda_l = 1\ for\ l > 1$ are fixed. $\gamma = 1$ is always considered as 1 for $l > 1$ because variation of this parameter does not indulge any change the separation performance.

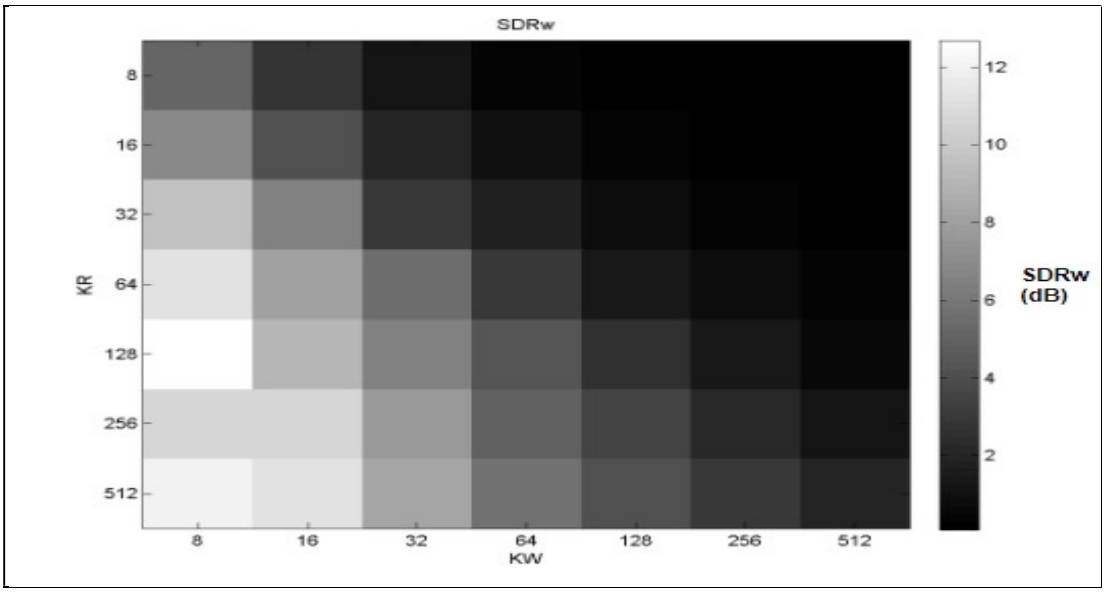


Figure 5.5.1 SDRw results obtained from optimisation process for the proposed ST-NMPCF approach with varying K_R and K_W parameters by fixing the $\lambda_1 = 0.001$ and $\gamma = 1$

From the SDR_W results depicted in ***Figure 5.5.1*** we get a lot of information to discuss. The first thing to be noticed is we get the best estimated wheezing signal at $K_R = 128$ and $K_W = 8$ and its $SDR_W = 12.9dB$. The worst performance has the $SDR_W = 1.9dB$. ***Figure 5.5.1*** depicts that the performance of the algorithm is good if the components of $K_R > K_W$. The algorithm is almost removing interfered respiratory components from the mixture signal if the allocated number of components to common or respiratory basis vectors K_R are greater than individual or wheezing component vectors KW. This is because by allocating the more number of respiratory basis vectors the magnitude spectrogram of mixture signal X is properly partially co-factoring both the spectral features of respiratory components obtained from training signal Y and temporal features from mixture signal X. As a result the mixture signal is properly able to remove the respiratory sounds leaving the wheezing sounds behind with no respiratory sounds left. On emore observation that we obtained from Figure Also Within the parameter space of $K_R = [64 - 512]$ and $K_R = [8 - 32]$the overall separation performance of the signal is depicted to be good with SDR_W values ranging between 12.8 dB to 11.15 dB allowing a minimum of 2dB loss for best separation performance. Moreover the overall separation performance of the given algorithm is worst within the parameter space

of $K_R = [8 - 32]$ and $K_R = [64 - 512]]$. The signal which obtained best performance from the database D1 is the 46th mixture signal named as 'M46_0dB.wav' in the database.

We also examined the in separation performance of the proposed ST-NMPCF approach for the estimated respiratory signal $\hat{x}_r[n]$, we evaluated the values by varying the parameters as mentioned in the ***Table 5.5.2*** and ***Figure 5.5.2*** depicts the SDR_R values obtained by varying K_R and $K_W$values and fixing the other parameters with their optimal values as $\lambda_1 = 0.001$ and $\lambda_l = 1\ for\ l > 1$ and $\gamma = 1$. Form the figure the highest SDR_R value is noted as 0.13 dB and the worst value is noted as -1.7 dB

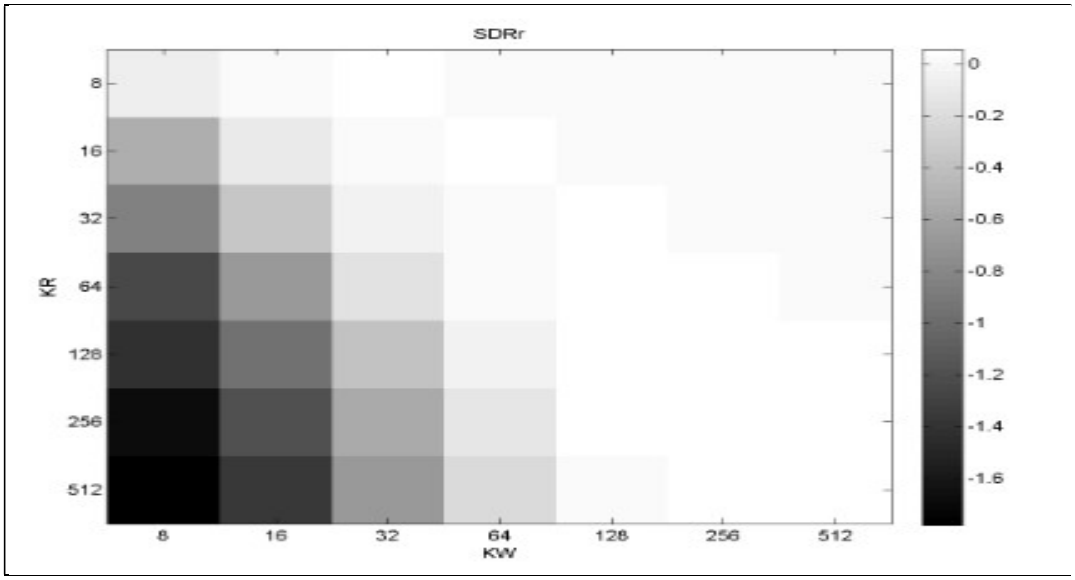
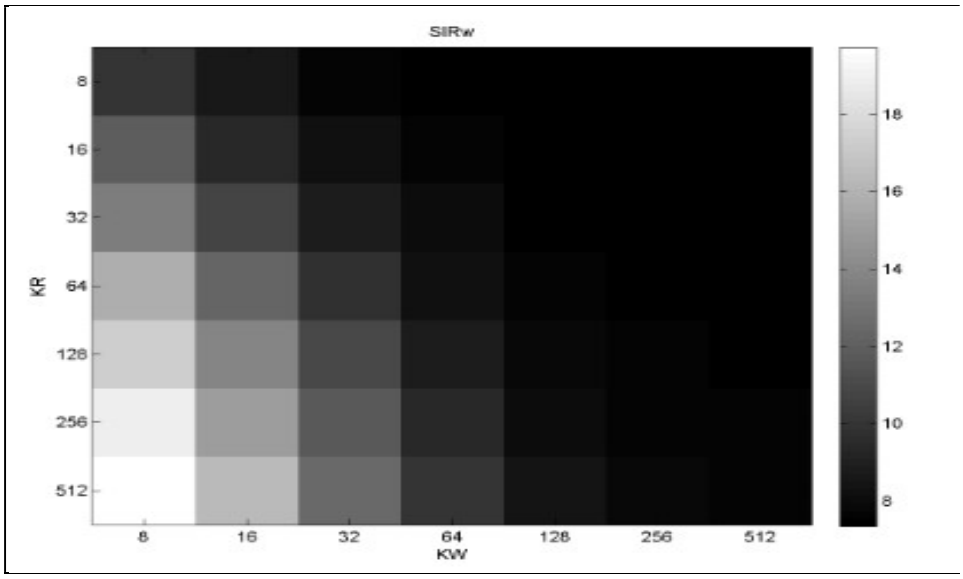


Figure 5.5.2 SDRr results obtained from optimisation process for the proposed ST-NMPCF approach with varying K_R and K_W parameters by fixing the $\lambda_1 = 0.001$ and $\gamma = 1$.

To study about the interferences of respiratory sound in estiamted wheeze signal $\hat{x}_w[n]$, and interference of wheeze sound in separated respiratory signal $\hat{x}_r[n]$, We evaluated the SIR metric for the estimated signals. The results obtained are depicted in ***Figure 5.5.3***, It depicts the obtained results by varying the same parameters as defined in the ***Table 5.5.2. Figure 5.5.3 A*** depicts the SIRw values and ***Figure 5.5.3 B*** depicts the SIRr values.

A



B

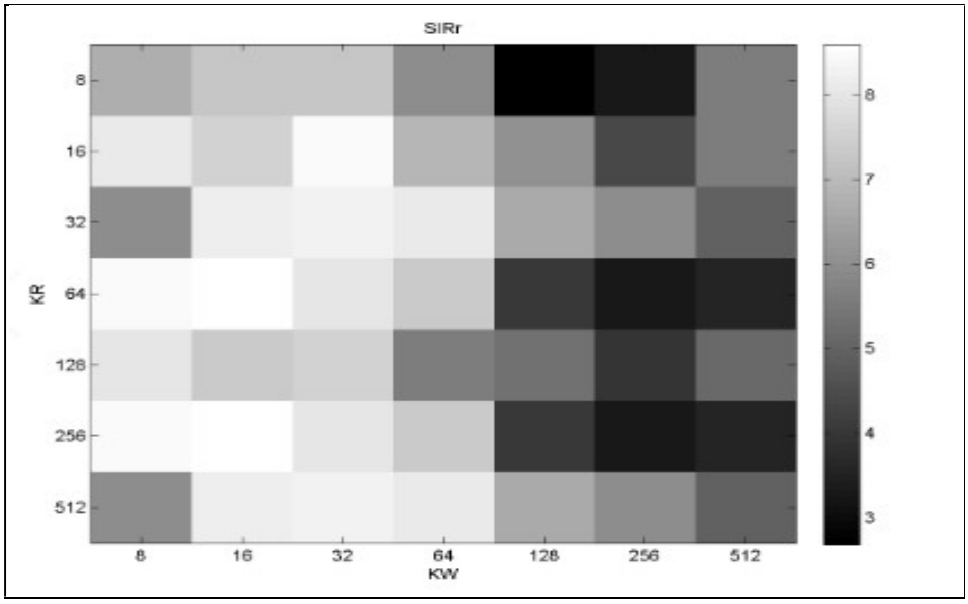


Figure 5.5.3 SIR results obtained from optimisation process for the proposed ST-NMPCF approach with varying K_R and K_W parameters by fixing the $\lambda_1 = 0.001$ and $\gamma = 1$ .Note A represents SIRw results and B represents SIRr results

By analysing the ***Figure 5.5.3*** the the best value obtained for SIRw= 19.7 dB and SIRr = 8.8 dB. From ***Figure 5.5.3*** **A** we can notice that the interference of respiratory components in estimated wheezing signal $\hat{x}_w[n]$, are increasing for parameter values of $K_R > K_W$.This is because by increasing the components of the respiratory sounds the common basis matrix A_R is collecting all the interfered respiratory sounds which is a sign of no or very low intensity respiratory sounds in the estimated

wheezing signal $\hat{x}_w[n]$. Therefore to obtain best performance with very less interference of respiratory sounds in the estimated wheezing signal the common basis vectors of respiratory components must be taken greater than the individual wheezing components. Also from analysing ***Figure 5.5.3 B*** we can conclude that the interference of wheezing sound in estimated respiratory signal $\hat{x}_r[n]$ for most of the parameter combinations is not much.

Finally from all the discussion so far the main points to conclude are: the best separated wheezing signal have the metric values as SDR_W=12.8 dB SIR_W=19.7dB and SAR_W=13.5 dB

Separation results obtained from optimisation process for proposed approach:

In this particular block of draft we are going to show the separation performance results obtained through the spectrogram representation of separated signals along with the mixture signal tested. ***Figure 5.5.4*** depicts the tested respiratory mixture signal $x[n]$ which provided the best separation performance results from database D1 obtained during optimisation process. The colour bar in ***Figure 5.5.4*** represents the various intensities of sounds present in the provided mixture signal. By observation we conclude that the wheezing sounds are with more intensity than respiratory sounds that implies these sounds are louder than the underlying breath sounds. We had already mentioned this point in section 2.4.2 that the wheezes are usually heard more louder than the respiratory sounds.

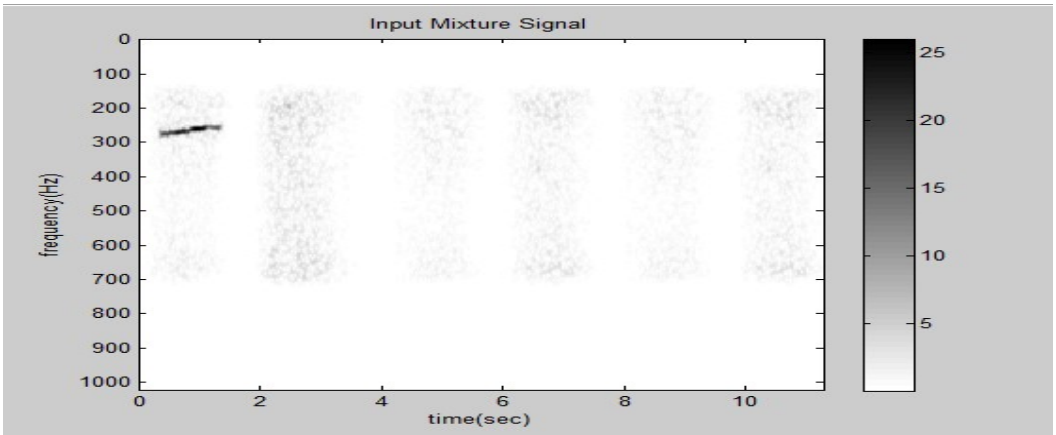


Figure 5.5.4 Spectrogram representation of the tested respiratory mixture signal that provided best performance in optimisation process.

Figure 5.5.5 depicts the spectrogram of estimated respiratory signal $\hat{X}_R$ that is obtained through evaluating the proposed ST-NMPCF approach applied on mixture signal X. From the colour bar we can conclude that the most of the indulged respiratory sounds in the provided mixture signal are ranging between 4.7 dB to 0.1dB and are removed from the mixture signal.

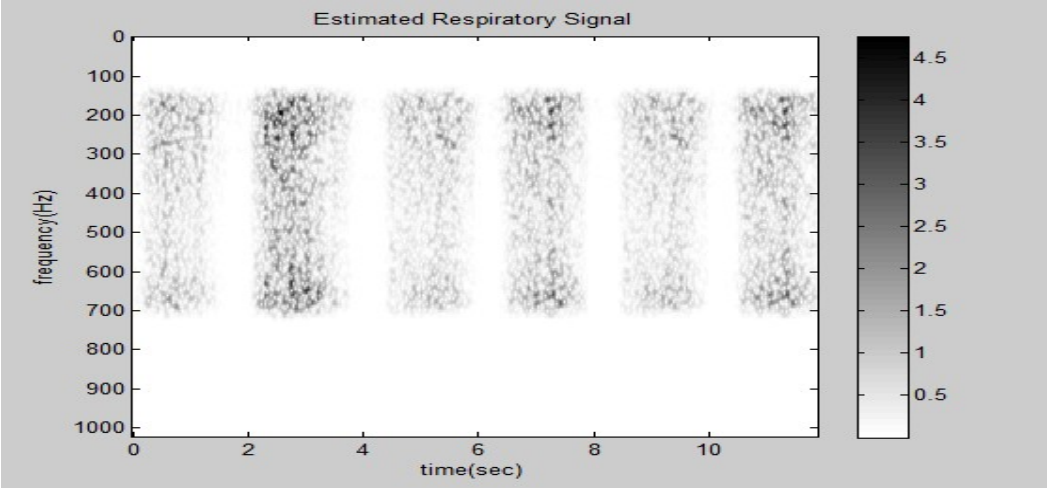


Figure 5.5.5 Spectrogram of the estimated respiratory signal $\hat{X}_R$ obtained by testing the mixture signal X through the proposed ST-NMPCF approach.

Similarly ***Figure 5.5.6*** depicts the spectrogram representation of the separated or estimated wheezing signal $\hat{X}_W$.

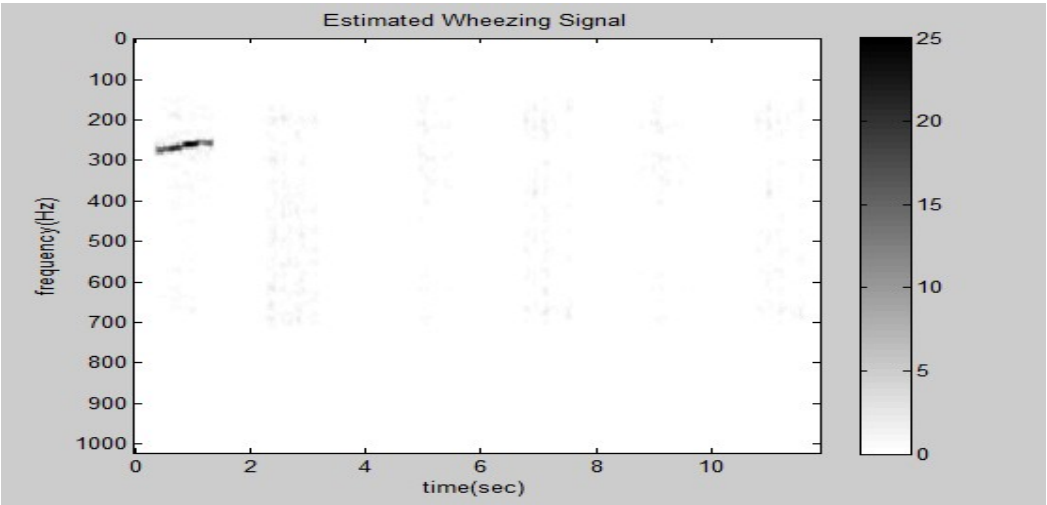


Figure 5.5.6 Spectrogram of the estimated wheezing signal $\hat{X}_W$ obtained by testing the mixture signal X through the proposed ST-NMPCF approach.

From the colour bar in ***Figure 5.5.6*** we can observe that the spectrogram contains sounds of higher intensity (wheezing sounds) and also we can conclude that most of the respiratory sounds are removed except the very low intensity respiratory sounds. Therefore with our proposed approach we can acquire good audio quality wheezing sounds from the respiratory mixture signal.

Separation results obtained for S-NMPCF T-NMPCF and ST-NMPCF approach

Now we are going to explain how proposed ST-NMPCF obtained good results compared to S-NMPCF and T-NMPCF approaches. We tested the same mixture signal ***Figure 5.5.4*** that provided best results by running our proposed ST-NMPCF approach , S-NMPCF approach and T-NMPCF approach. ***Table*** represents the characteristics of different parameters considered for comparison of different source separation approaches.

Characteristics	S-NMPCF	T-NMPCF	ST-NMPCF
Training signal	$y[n]$	_	$y[n]$
Respiratory mixture signal	$x[n]$	$x[n]$	$x[n]$
Segmentation	_	l-column blocks $1 < l < L$ AMIE_SEG [95]	l -column blocks $1 < l < L$ AMIE_SEG [95]
Parameters considered	K_R, K_W,λ	K_R, K_W	K_R, K_W,λ

Table 5.5.3 Parameter characteristics of compared source separation approaches

Figure 5.5.7 A, ***Figure 5.5.7 B*** and ***Figure 5.5.7 C*** represents the spectrogram representations of estimated wheezing signal obtained from S-NMPCF approach, T-NMPCF approach, ST-NMPCF approach for the mixture signal in ***Figure 5.5.4***

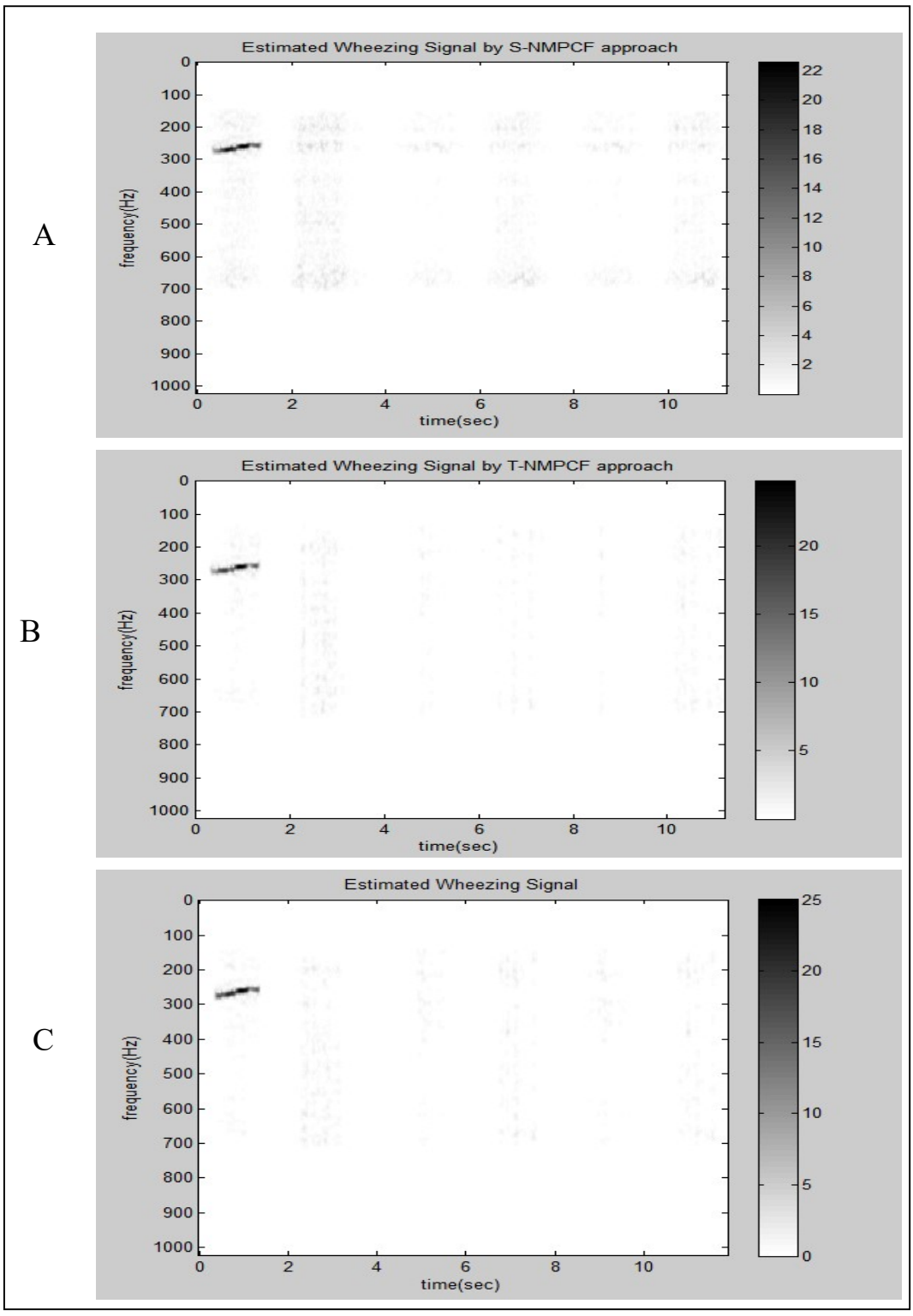


Figure 5.5.7 Spectrogram representations of estimated wheezing signals obtained by A) S-NMPCF approach, B) T-NMPCF approach and C) ST-NMPCF approach

By observing spectrogram representations of separated wheezing sound signal by three source separation approaches S-NMPCF, T-NMPCF and ST-NMPCF approach from the ***Figure 5.5.7 A, Figure 5.5.7 B and Figure 5.5.7 C*** we conclude that proposed ST-NMPCF approach resulted in good wheezing sound separation than rest of the two. It almost removed respiratory sound components. But in the other two approaches we can see there are still some respiratory components left in the separated wheezing signal. We concluded this point by observing the colour bar of the spectrograms of estimated wheezing signal. The colour bar of the ST-NMPCF approach showed higher intensity sounds than S-NMPCF approach and the higher intensity sounds of estimated wheezing signal are less than the T-NMPCF approach. Though the evaluated performance of T-NMPCF results are worse than the S-NMPCF approach. But the T-NMPCF approach widens the case of the NMPCF system where the respiratory database is not available. Admittedly, the temporally repeating property is not enough to cover all the different spectrums of respiratory sources, either. But the ST-NMPCF approach by utilising the advantages of both S-NMPCF approach and T-NMPCF approach removed most of the interfered respiratory components. Therefore we can conclude one point from these observations i.e. source separation approach considering prior knowledge or training signal provided good separation results (S-NMPCF and T-NMPCF) than the one without considering the training signal(T-NMPCF). This is because while considering the training signal the estimated matrices try to make use of spectral information of the training respiratory signal along which are common in all respiratory sounds.

Figure 5.5.8 A, ***Figure 5.5.8 B*** and ***Figure 5.5.8 C*** represents the spectrogram representations of estimated respiratory signal obtained from S-NMPCF approach, T-NMPCF approach, ST-NMPCF approach for the mixture signal in ***Figure 5.5.4***

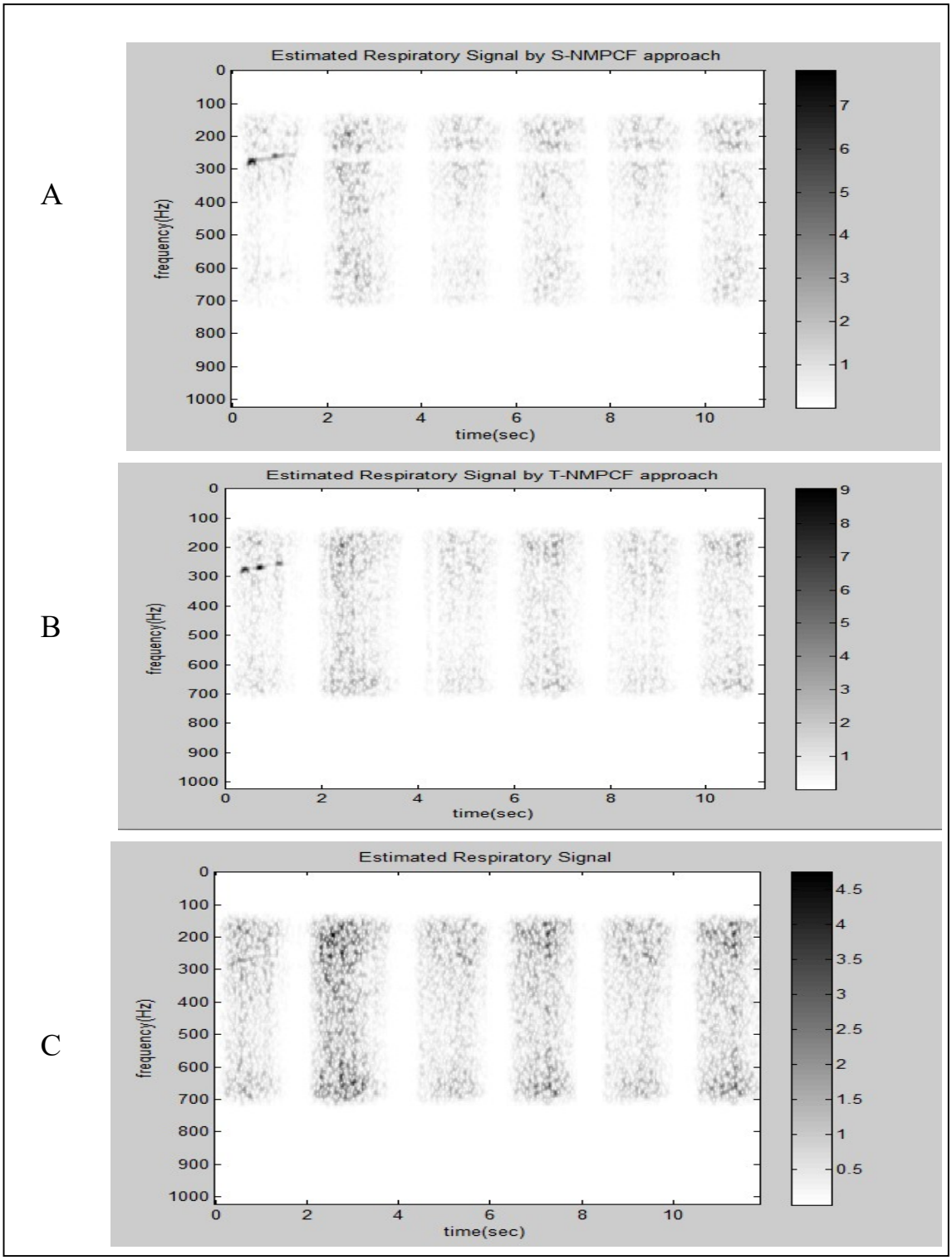


Figure 5.5.8 Spectrogram representations of estimated respiratory signals obtained by A) S-NMPCF approach, B) T-NMPCF approach and C) ST-NMPCF approach

5.6 Testing

In this particular section of the draft we are going to explain when the proposed source separation approach works with best results and worst results and what characteristics of the respiratory mixture signal provide the best results. In order to conclude various aspects we used database D2 to evaluate the results. While evaluating the signals from testing dataset D2 We observed few things i.e the mixture signal with louder wheezing sound than respiratory signal provided good results and the respiratory signal with louder respiratory sound than wheezing signal provided very worse results.

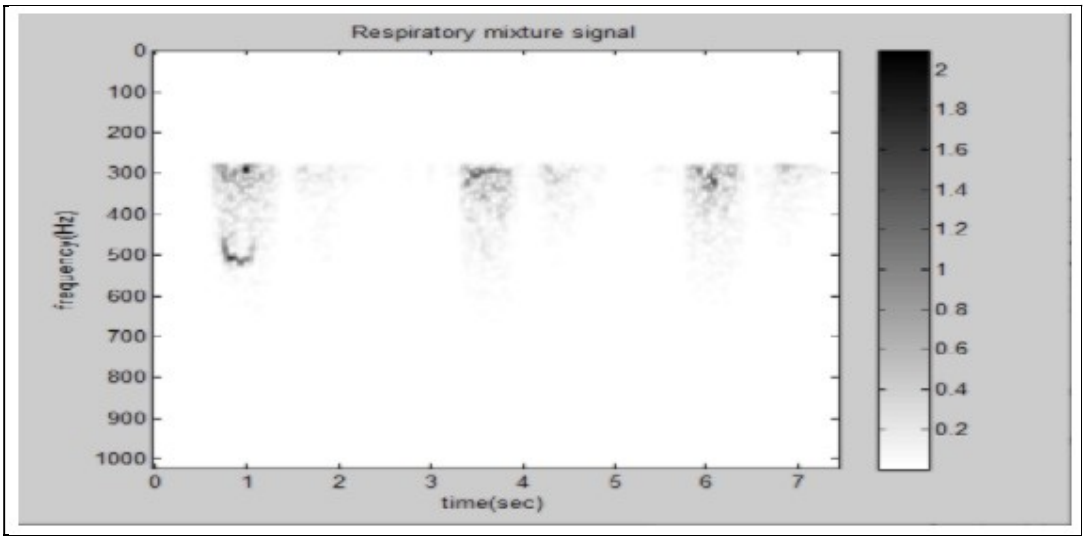


Figure 5.6.1 Respiratory mixture signal from Database D2 which provided worst performance

Figure 5.7.1 represents the spectrogram of the respiratory mixture signal from which we obtained worst performance. We evaluated the intensity of the respiratory sounds and wheezing sounds Using Matlab and we found that the intensity of some of the respiratory components is higher than wheezing components. By evaluating SDR_W results of the input mixture signal depicted in ***Figure 5.7.1*** by varying various parameters when we tried to analyse the estimated wheezing signal, the result obtained as depicted in ***Figure 5.7.2***

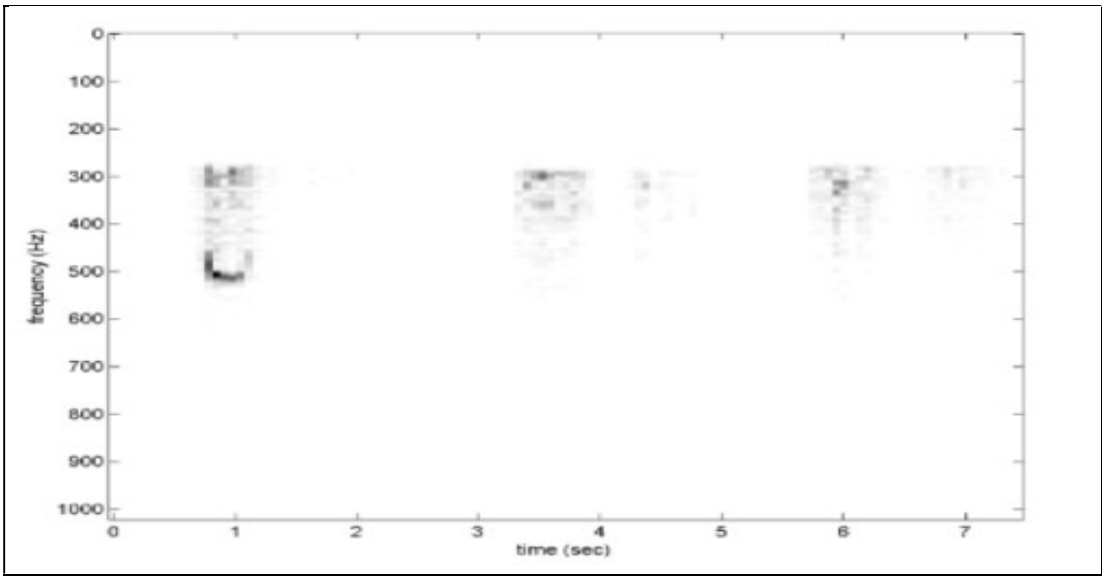


Figure 5.6.2 Spectrogram representation of the worst estimated wheezing signal obtained by running database D2

From observation we concluded that the high interference of the respiratory sounds in the mixture signal which diminish the sound of wheezing sound in the mixture signal made the source separation technique work at a satisfactory rate. From ***Figure 5.7.2*** although wheezing sound got separated but there are lots of respiratory components involved in the separated wheeze signal which lowered the performance criteria. The database D2 used in testing process is evaluated by considering the same parameters mentioned in ***Table 5.5.1***

The Best performance results obtained while evaluating dataset D2 are depicted in ***Figure 5.7.3***. The main characteristic to note about this signal is the loudness of wheezing sound is higher than respiratory sounds. Therefore by analysing the results obtained we would like to conclude that if the loudness of the interfered respiratory sound is less than the wheezing sound the source separation performance is good.

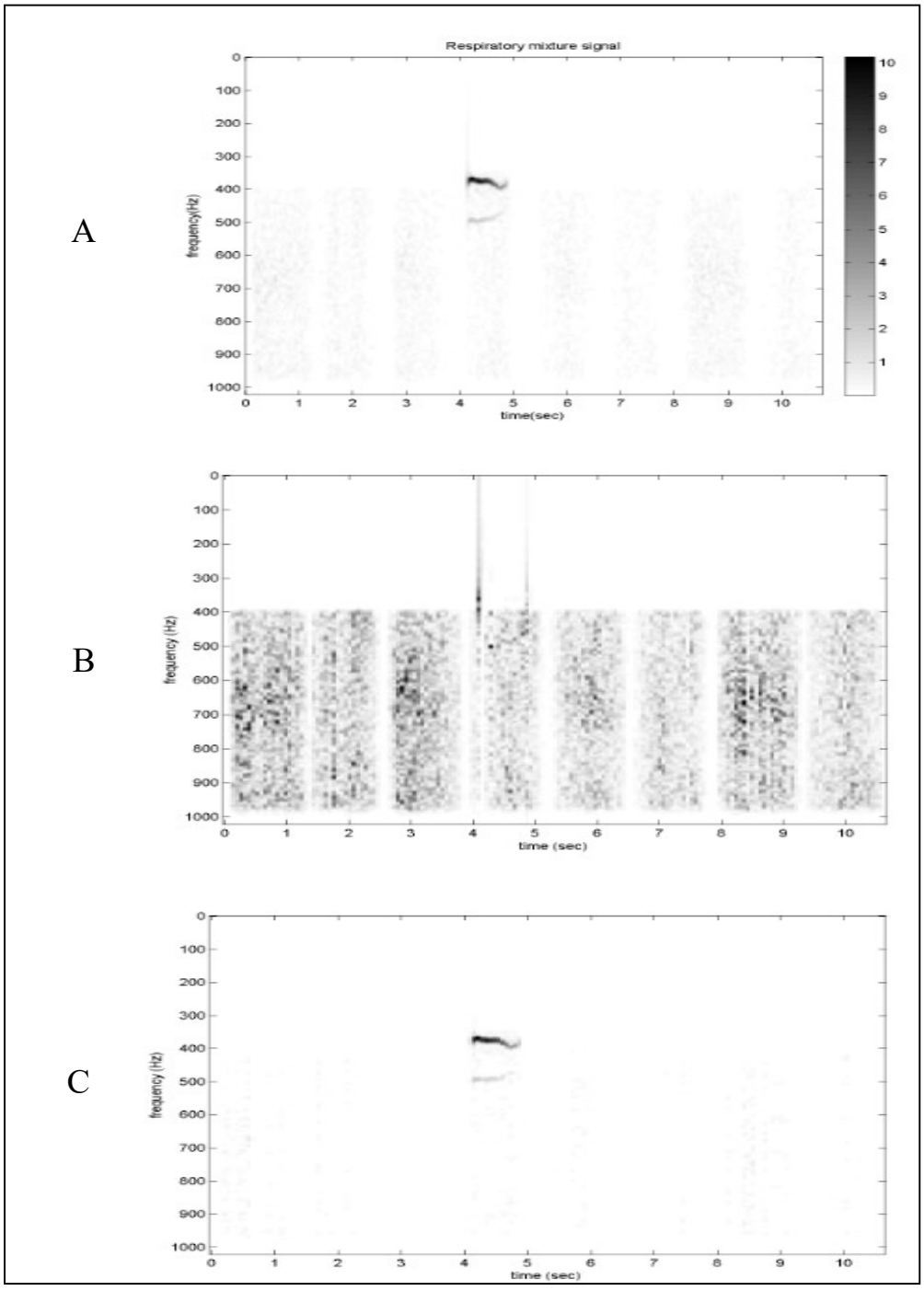


Figure 5.6.3 Spectrogram representation input respiratory mixture signal (A), estimated respiratory signal (B), estimated wheezing sound (C) of best performance results obtained by running database D2

There is also another criteria where the performance results are worse. This is when there are repetitive wheezing sounds in the segmented blocks. For the mixture signal in ***Figure 5.7.4*** we added the same wheezing sound components in all the respiratory events to test how the

source separation performance is working. The results we obtained are depicted below.

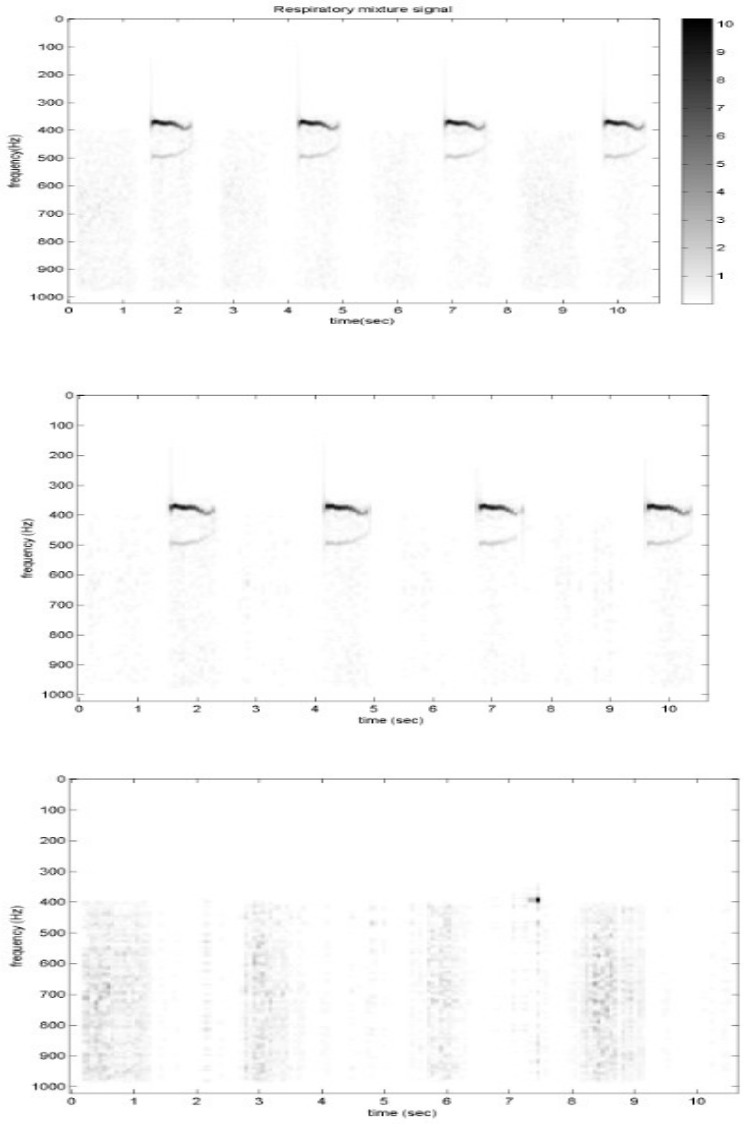


Figure 5.6.4 Spectrogram representation of input respiratory mixture signal (Top), estimated respiratory signal (middle) , estimated wheezing signal (bottom) when wheezing components are included in all the respiratory events.

To prove that the proposed ST-NMPCF approach is a better source separation technique than S-NMPCF and T-NMPCF approach, we computed the SDR, SIR, SAR values of estimated wheezing sounds $\hat{x}_w[n]$ and estimated respiratory sounds $\hat{x}_r[n]$ obtained from all the three

approaches by running all the 16 files from database D2 and we obtained the box plots as depicted in ***Figure 5.6.5,Figure 5.6.7, Figure 5.6.8.***

Generally higher values of SDR, SIR, SAR depicts the higher quality of estimated sounds. From observing ***Figure 5.6.5*** the SDR_W and SDR_R values of S-NMPCF approach are having higher values compared to our proposed ST-NMPCF approach

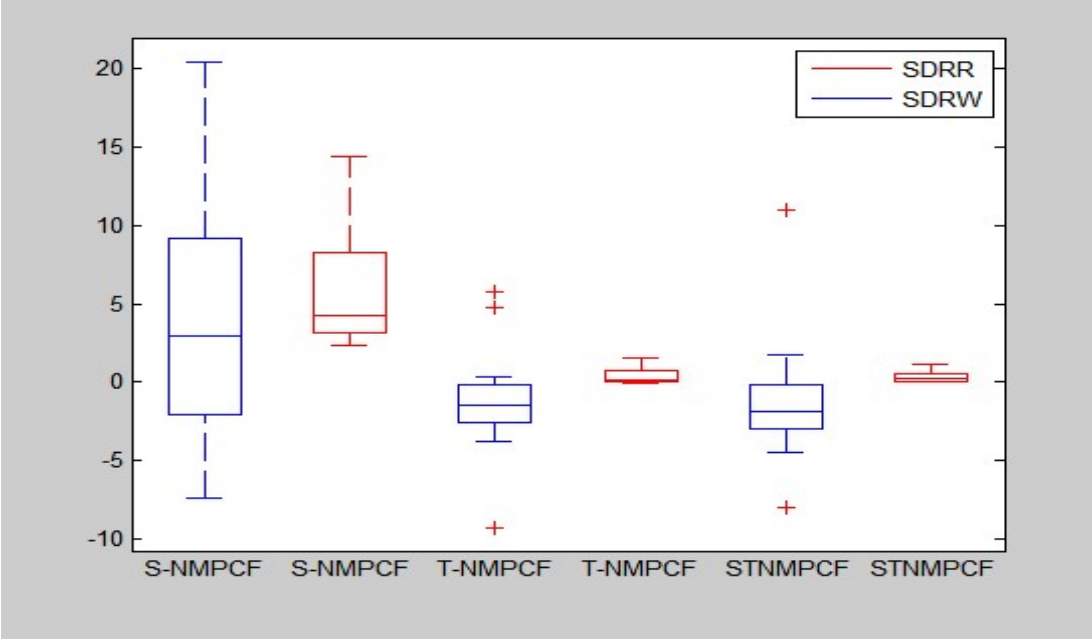


Figure 5.6.5 SDRW and SDRR results evaluating the database D2 . Note SDRW results represented in blue boxes and SDRR results represented in red boxes

From equation (5.2.3) SDR computes the energy ratio of estimated signal to the error terms ($e^e_{interf}[n], e^e_{spatial}[n], e^e_{artifacts}[n]$). This implies higer values of SDR depict that the error terms have low value and vice-versa. From ***Figure 5.6.5*** the maximum SDR_W value that we obtained by running database D2 for ST-NMPCF approach is around 12.5 dB (the highest data point value). Whereas for S-NMPCF approach the highest value is around 21dB. As well as most of the SDR_W values of ST-NMPCF approach are ranging between 0 dB to -3dB. Therefore according to our obtained SDR values, main point to conclude S-NMPCF algorithm provides good audio quality estimated signals with low value errored terms than our proposed ST-NMPCF algorithm. Though S-NMPCF approach provides good quality audio signals than ST-NMPCF approach but we don't know about the interference of respiratory sounds in the estimated wheezing signals obtained from S-

NMPCF, ST-NMPCF and T-NMPCF approaches. To conclude about the interference of respiratory sounds in the estimated signals we have to examine the SIR metric values of estimated signals obtained by running these three approaches. ***Figure 5.6.6*** depicts the box plots of SIR values of estimated signals obtained from S-NMPCF, T-NMPCF and ST-NMPCF approaches by running database D2.

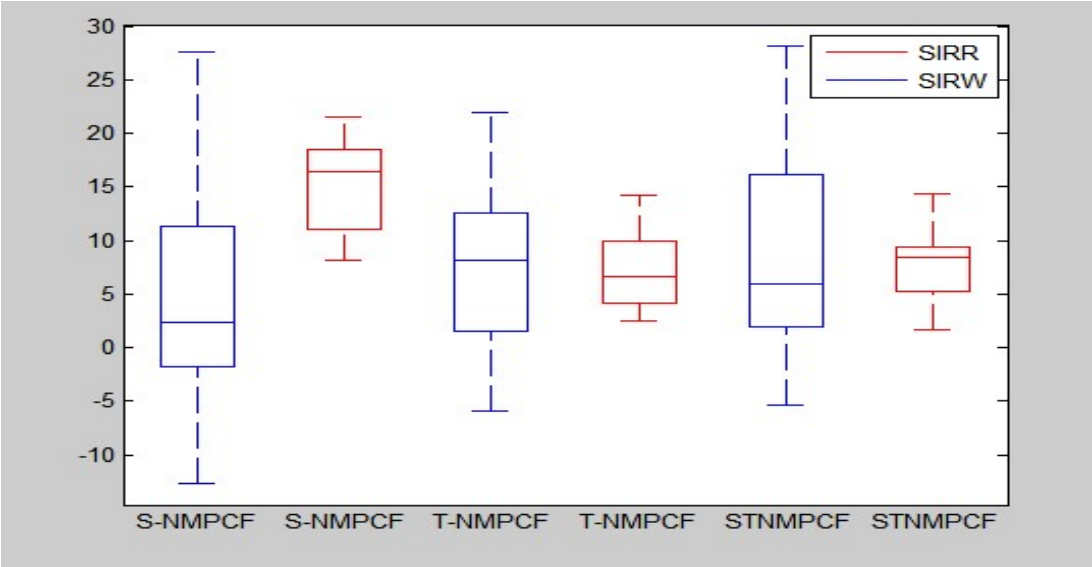


Figure 5.6.6 SIRW and SIRR results evaluating the database D2 . Note SIRW results are represented in blue boxes and SIRR results are represented in red boxes

According to the definition of SIR from equation (5.2.4) SIR computes the energy ratio of estimated signal to the error term of interference of other sources ($e^{e}_{interf}[n]$) in the estimated signal. This implies, higher SIR values indicate that the interference of other sources in estimated signal are less and vice-versa. Therefore SIR_W values indicate about interference of respiratory sounds in the estimated wheezing signals. If we observe ***Figure 5.6.6*** the SIR_W values of proposed ST-NMPCF approach are indicating higher values than the S-NMPCF and T-NMPCF approaches. The SIR_W values of the ST-NMPCF are ranging with a highest value of 29dB to a least of -4dB and most of the SIR_W values evaluated by database D2 are between 17dB to 3dB with a median of 5dB. Whereas for S-NMPCF approach the SIR_W values are ranging with the highest value of 27dB to the least of -14dB and for most of the estimated wheezing signals from database D2 the SIR_Wvalues are ranging between 11dB to -2dB. For T-NMPCF approach the

SIR_W values are ranging with the highest value of 20dB to the least of -6dB and for most of the estimated wheezing signals from database D2 the SIR_W values are ranging between 14dB to 2dB. Therefore by observing the SIR_W values of three NMPCF approaches what we can conclude is that ST-NMPCF algorithm is removing most of the interfered respiratory sounds from the mixture signal giving more importance to wheezing sounds in the estimated wheezing signal. It indicates that a more reliable modelling of respiratory sounds is achieved using jointly the shared respiratory spectral patterns along the segments and a prior knowledge of the respiratory spectral content by means of the respiratory training signal. Besides ST-NMPCF the SIR_W values are good for S-NMPCF approach than T-NMPCF approach. Therefore we can conclude that use of training signal improves the segregation performance by removing most of the spectral content of respiratory sounds from mixture signal than without training signal as in T-NMPCF.

Also, the SDR_W and SIR_W values of ST-NMPCF, S-NMPCF and T-NMPCF approach from ***Figure 5.6.5*** and ***Figure 5.6.6*** indicate that though the SDR_W values of proposed ST-NMPCF approach are low but the proposed algorithm is removing most of the interfered respiratory sounds from the mixture signal providing only the wheezing sounds in the estimated signal. Whereas from SDR_W and SIR_W values of S-NMPCF approach we can conclude that this algorithm provides good quality signals but the interference of respiratory sounds still occur in estimated wheezing signal which is not a good choice. Whereas the performance factor of T-NMPCF is low compared to S-NMPCF and ST-NMPCF approach which implies blind source separation approach though separate the target sources from mixture signal , they separation efficiency is not better than other two approaches.

There is another metric called SAR (Source to Artifacts Ratio) which computes the interference of artifacts that occur during execution of algorithm. The SAR metric is calculated by using the equation (5.2.5). ***Figure 5.6.7*** depicts the SAR values of estimated respiratory sounds and estimated wheezing sounds for S-NMPCF, T-NMPCF and ST-NMPCF approaches obtained by evaluating database D2.

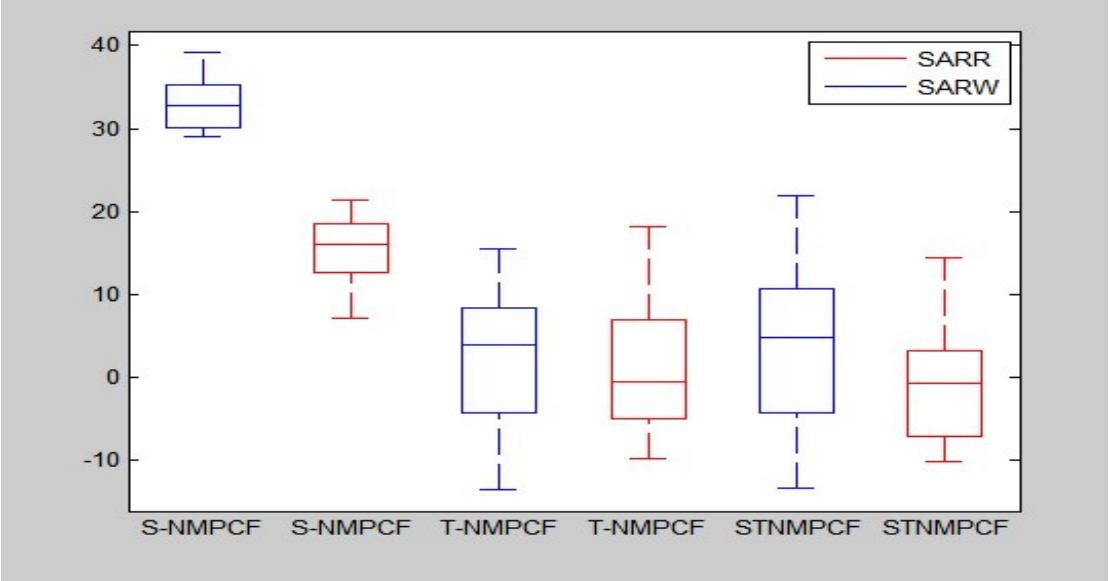


Figure 5.6.7 SARW and SARR results evaluating the database D2 . Note SARW results represented in blue boxes and SARR results represented in red boxes.

From ***Figure 5.6.7*** the best and worst SAR_Wvalues obtained for ST-NMPCF approach are 21dB and -14dB, whereas the best and worst SAR_Wvalues obtained for S-NMPCF approach are 39dB and 29dB and the best and worst SAR_Wvalues obtained for ST-NMPCF approach are 12dB and -14dB. This implies SAR_W values for S-NMPCF approach are high compared to the values of T-NMPCF and ST-NMPCF approach. This implies proposed ST-NMPCF and T-NMPCF approach induces more artifacts during program execution than S-NMPCF approach.

Finally to conclude from the observations done so far, ST-NMPCF approach provides the best estimated wheezing signal with very less interference of respiratory sounds at an expense of decreasing the audio quality. Whereas S-NMPCF approach provides estimated wheezing signals with good audio quality but at an expense of interfered respiratory sounds in the estimated wheezing signal. Contrastingly, the separation performance of T-NMPCF is worst compared to supervised S-NMPCF and ST-NMPCF approaches. Therefore, according to our observations proposed ST-NMPCF approach is the best segregation approach to consider, although it induces artifacts and provides low SDR values because this approach completely removes the unwanted respiratory sounds from the mixture signal than compared to S-NMPCF and T-NMPCF approaches. As main aim our thesis statement is to

separate the wheezing sounds with no remnants of respiratory sounds in them which can be accomplished well with proposed ST-NMPCF approach than S-NMPF and T-NMPCF approach. This is because we can improve the audio quality of estimated signals and can remove the artifacts in them by adding some constraints with the guide of some research papers.

6 DISCUSSION AND CONCLUSION

In this particular section of the draft we are going to discuss about the accomplished objectives that we have mentioned in chapter 3 and conclude about important observations and results. First and foremost objective of our thesis study is to study about various reference studies related to main topic of our master thesis topic, which we have discussed in state of the art methods (section 2.5). In this section we have mentioned about widely used approaches to detect and separate wheezes from mixture signal along with explanations of NMF, spectral trajectories and few others. Our second objective is to design an algorithm in MATLAB to segregate wheezing sounds from respiratory audio mixtures the explanation about which is discussed briefly in section 4.3.4. Creation of databases composed of wheezes and respiratory sounds from various internet pulmonary repositories is one of the important objectives of our master thesis study. Section 5.1 explains about our created databases D1, D2 and D3 composed of respiratory and wheezing sounds. Our next objective is to evaluate the performance of the implemented algorithm, which we accomplished through optimisation and testing process explained in section 5.5 and section 5.6. Creation of user friendly interface in MATLAB for the user to make the evaluation process user-friendly is another important objective of our thesis study, which we explained briefly in Appendix (chapter 9). Therefore from this discussion we conclude that we tried to achieve all the objectives that have to be done during our thesis study.

Also if we mention about some of the conclusions that we observed during our master thesis study are:

1. Supervised approaches showed better separation performance than the unsupervised approaches. This point can be demonstrated from results obtained in testing (section 5.6) and optimisation process (section 5.5). This is because supervised approaches use respiratory training signal which is helpful in modelling the spectral components of respiratory sounds during partial co-factorisation process as a result provides good estimated wheezing

signals by removing most of the spectral content of the respiratory sounds from the mixture signal

2. From the SDR_W, SIR_W and SAR_W results explained in section 5.6 S-NMPCF approach has $SDR_W = 21dB$, $SIR_W = 26dB$ and $SAR_W = 38dB$. Whereas T-NMPCF approach has $SDR_W = 6dB$, $SIR_W = 22dB$ and $SAR_W = 8dB$ and proposed ST-NMPCF approach has $SDR_W = 12.5dB$, $SIR_W = 29.5dB$ and $SAR_W = 12dB$. From our results we want to conclude that ST-NMPCF approach is best separation approach for getting best estimated wheezing signals with most of the respiratory sounds removed because of high SIR_W value, but at an expense of losing the audio quality because of low SDR_W value compared to S-NMPCF approach. Though S-NMPCF approach provide good audio quality wheezing sounds, the respiratory sounds still remain in estimated wheezing signal which is not a good choice.
3. The evaluation results may vary from person to person based on the respiratory and wheezing samples they collect from internet. We were not able to get enough number of good respiratory audio mixture signals which resulted in poor audio quality of proposed approach but our proposed algorithm worked well in almost removing the interfered respiratory sounds from mixture signal.
4. The main drawback of working on wheeze separation approaches is having limited number of publicly available pulmonary samples database to evaluate the study. The researchers who make research on respiratory sound analysis i.e. for detection, separation or any sort of research related to similar topic they have to create their own respiratory sound database which contains solo-respiratory and solo-wheezing sounds. With them they can mix the respective respirator sounds and wheezing sounds to form a mixture signal which is a time-consuming and tiresome task.

7 FUTURE PLANS

In our future works we want improve the audio quality of the estimated wheezing signal for the proposed ST-NMPCF approach by increasing the SDR and SAR values. We would also work on increasing robustness of the propsed ST-NMPCF approaches by adding various constraints like sparsness and smoothness to the input mixture signal. From our observations, it is known that too long respiratory training signal does not improve the separation quality, because incongruent parts can harm the reconstruction. Therefore in our future works we would like to research on relationship between the size of training signal used for S-NMPCF or ST-NMPCF used in separation performance. In our future work will focus on the introduction of new constraints to the NMPCF approaches to model wheezing sounds according to their spectral content in order to automatically distinguish the severity of the lung disorder.

8 REFERENCES

1. World Health Organization. Asthma Fact Sheet; 2017. Available from: http://www.who.int/mediacentre/factsheets/fs307/en/.

2. World Health Organization. Global surveillance, prevention and control of chronic respiratory diseases. A comprehensive approach. Geneva, WHO, 2007. Available from: https://www.who.int/gard/publications/GARD_Manual/en/.

3. World Health Organization. Chronic obstructive pulmonary disease (COPD) Fact Sheet; 2017. Available from: http://www.who.int/mediacentre/factsheets/fs315/en/.

4. World Health Organization. World Health Statistics 2008; Available from: http://www.who.int/gho/publications/world_health_statistics/.

5. Trueman D, Woodcock V, Hancock E. Estimating the economic burden of respiratory illness in the UK; 2017. Available from: https://www.blf.org.uk/what-we-do/our-research/economic-burden.

6. Balachandran J, Teodorescu M. Sleep Problems in Asthma and COPD. Am J Respir Crit Care Med. 2013;. [Google Scholar].

7. National Health Service England. Overview of potential to reduce lives lost from Chronic Obstructive Pulmonary Disease (COPD); 2014. Available from: https://www.england.nhs.uk/wp-content/uploads/2014/02/rm-fs-6.pdf.

8. British Thoracic Society and Scottish Intercollegiate Guidelines Network. British guideline for the management of asthma; a national clinical guideline (SIGN 153); 2016. Available from: https://www.brit-thoracic.org.uk/guidelines-and-quality-standards/asthma-guideline/.

9. British Thoracic Society and Scottish Intercollegiate Guidelines Network. Chronic obstructive pulmonary disease in over 16s: diagnosis and management; 2010. Available from: https://www.nice.org.uk/guidance/CG101.

10. Burrows B., Huang N., Hughes R., Johnston R., Kilburn K., Kuhn C., Miller W., Mitchell M., Snider G. Pulmonary Terms and Symbols. Chest. 1975;67:583–593. doi: 10.1378/chest.67.5.583. [CrossRef] [Google Scholar].

11. Taplidou S.A., Hadjileontiadis L.J. Wheeze detection based on time frequency analysis of breath sounds. Comput. Biol. Med. 2007;37:1073–1083. doi: 10.1016/j.compbiomed.2006.09.007. [PubMed] [CrossRef] [Google Scholar].

12. Shaharum S.M., Sundaraj K., Palaniappan R. A survey on automated wheeze detection systems for asthmatic patients. Bosn. J. Basic. Med. Sci. 2012;12:249–255. [PMC free article] [PubMed] [Google Scholar].

13. Jacome C., Marques A. Computerized Respiratory Sounds in Patients with COPD: A Systematic Review. COPD. 2015;12:104–112. doi: 10.3109/15412555.2014.908832. [PubMed] [CrossRef] [Google Scholar].

14. Gurung A., Scrafford C.G., Tielsch J.M., Levine O.S., Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis. Respir. Med. 2011;105:1396–1403. doi: 10.1016/j.rmed.2011.05.007. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

15. Aras S, Ozturk M, Gangal A. Endpoint detection of lung sounds for electronic auscultation. In: IEEE 2016 ¨ International Conference on Telecommunications and Signal Processing; 27–29 June 2016; Vienna, Austria. New York, NY, USA: IEEE. pp. 405-408.

16. Emmanouilidou, D., McCollum, E. D., Park, D. E., & Elhilali, M. (2015). Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries. *IEEE Transactions on Biomedical Engineering*, 62(9), 2279-2288. [Google Scholar].

17. Haider, Nishi Shahnaj, et al. Savitzky-Golay filter for denoising lung sound. *Brazilian Archives of Biology and Technology,* 2018, vol. 61. [Google Scholar].

18. Pourazad, M. T.; Moussavi, Z.; Thomas, G. Heart sound cancellation from lung sound recordings using time-frequency filtering. Medical and biological engineering and computing, 2006, vol. 44, no 3, p. 216-225. [Google Scholar].

19. Pourazad, M. T.; Mousavi, Z. K.; Thomas, G. Heart sound cancellation from lung sound recordings using adaptive threshold and 2D interpolation in time-frequency domain. *En Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE Cat. No. 03CH37439). IEEE, 2003. p. 2586-2589. [Google Scholar].

20. Taplidou, Styliani A., and Leontios J. Hadjileontiadis. "Wheeze detection based on time-frequency analysis of breath sounds." Computers in biology and medicine 37.8 (2007): 1073-1083. [Google Scholar].

21. Fiz, J. A., Jané, R., Homs, A., Izquierdo, J., Garcia, M. A., & Morera, J. (2002). Detection of wheezing during maximal forced exhalation in patients with obstructed airways. Chest, 122(1), 186-191. [Google Scholar].

22. Riella, R. J., Nohama, P., & Maia, J. M. (2009). Method for automatic detection of wheezing in lung sounds. Brazilian Journal of Medical and Biological Research, 42(7), 674-684. [Google Scholar].

23. M. Bahoura, Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Computers in Biology and Medicine, 2009. 39(9): 824-843. [Google Scholar]

24. A. Kandaswamy, C. Kumar, R. Ramanathan, Neural classification of lung sounds using wavelet coefficients. Computers in Biology and Medicine, 2004. 34(6): 523-537. [Google Scholar]

25. Y. Liu, C. Zhang, Y. Peng, Neural classification of lung sounds using wavelet packet coefficients energy. Trends in Artificial Intelligence, 2006, pp. 278-287. [Google Scholar]

26. L. Pesu, et al. Wavelet packet based respiratory sound classification. Proceedings of the IEEE-SP International Symposium on issue Date: 18-21 Jun 1996: IEEE. [Google Scholar]

27. M. Bahoura, and C. Pelletier. Respiratory sounds classification using Gaussian mixture models. Engineering in Medicine and Biology Society, IEMBS 26th Annual International Conference of the IEEE, 2004. [Google Scholar]

28. R. Folland, et al. Comparison of neural network predictors in the classification of tracheal-bronchial breath sounds by respiratory auscultation. Artificial Intelligence in Medicine, 2004. 31(3): 211-220. [Google Scholar]

29. J. Chien, et al. Wheeze detection using cepstral analysis in gaussian mixture models. Engineering in Medicine and Biology Society, 2007, 29th Annual International Conference of the IEEE. [Google Scholar]

30. Torre-Cruz, J., Canada's-Quesada, F., Vera-Candeas, P., Montiel-Zafra, V., & Ruiz-Reyes, N. (2018, May). Wheezing sound separation based on constrained non-negative matrix factorization. *In Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology* (pp. 18-24). [Google Scholar]

31. Yoo, J., Kim, M., Kang, K., & Choi, S. (2010, March). Nonnegative matrix partial co-factorization for drum source separation. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1942-1945). IEEE. [Google Scholar].

32. Kim, M., Yoo, J., Kang, K., & Choi, S. (2010, March). Blind rhythmic source separation: Nonnegativity and repeatability. *In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 2006-2009). IEEE.* [Google Scholar]

33. Kim, M., Yoo, J., Kang, K., & Choi, S. (2011). Nonnegative matrix partial co-factorization for spectral and temporal drum source separation. *IEEE Journal of Selected Topics in Signal Processing* , 5 (6), 1192-1204.[Google Scholar]

34. Oud M, Doijes EH. Automated breath sound analysis. Proceedings of 18th Annual International Conference of IEEE Engineering in Medicine and Biology Society. Amsterdam: 1996. p 990-992. [Google Scholar]

35. Pasterkamp, Hans; Kraman, Steve S .; Wodicka, George R. Respiratory sounds: advances beyond the stethoscope. American journal of respiratory and critical care medicine , 1997, vol. 156, no 3, p. 974-987. [Google Scholar]

36. American Thoracic Society Ad Hoc Committee on Pulmonary Nomenclature. Updated nomenclature for membership reaction. ATS News. 1977;3:5–6. [Google Scholar]

37. Akasaka K., Konno K., Ono Y., Mue S., Abe C.Acoustical studies on respiratory sounds in asthmatic patients. Tohoku J. Exp. Med.1171975323333.

38. Information about Lung sounds Available online : Lung Sounds

39. Definition of Non-negative Matrix Factorisation Available online: Non-Negative Matrix Factorization definition

40. LEE, Daniel D.; SEUNG, H. Sebastian. Learning the parts of objects by non-negative matrix factorization. Nature, 1999, vol. 401, no 6755, p. 788-791. [Google Scholar]

41. FÉVOTTE, Cédric; VINCENT, Emmanuel; OZEROV, Alexey. Single-channel audio source separation with NMF: divergences, constraints and algorithms. En Audio Source Separation. Springer, Cham, 2018. p. 1-24. [Google Scholar]

42. GILLIS, Nicolas. The why and how of nonnegative matrix factorization. Regularization, optimization, kernels, and support vector machines, 2014, vol. 12, no 257, p. 257-291. [Google Scholar].

43. FÉVOTTE, Cédric; BERTIN, Nancy; DURRIEU, Jean-Louis. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. Neural computation, 2009, vol. 21, no 3, p. 793-830. [Google Scholar].

44. YANG, Zhirong, et al. Kullback-Leibler divergence for nonnegative matrix factorization. At the International Conference on Artificial Neural Networks . Springer, Berlin, Heidelberg, 2011. p. 250-257. [Google Scholar]

45. SMARAGDIS, Paris; RAJ, Bhiksha; SHASHANKA, Madhusudana. Supervised and semi-supervised separation of sounds from single-channel mixtures. In International Conference on Independent Component Analysis and Signal Separation . Springer, Berlin, Heidelberg, 2007. p. 414-421. [Google Scholar].

46. Supervised sound source separation using Nonnegative Matrix Factorization

47. VINCENT, Emmanuel; GRIBONVAL, Rémi; FÉVOTTE, Cédric. Performance measurement in blind audio source separation. IEEE transactions on audio, speech, and language processing, 2006, vol. 14, no 4, p. 1462-1469. [Google Scholar]

48. DIKMEN, Onur; CEMGIL, A. Taylan. Unsupervised single-channel source separation using Bayesian NMF. En 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. IEEE, 2009. p. 93-96. [Google Scholar]

49. Joder, C., Weninger, F., Eyben, F., Virette, D., & Schuller, B. (2012, March). Real-time speech separation by semi-supervised nonnegative matrix factorization. In International Conference on Latent Variable Analysis and Signal Separation (pp. 322-329). Springer, Berlin, Heidelberg. [Google Scholar]

50. LEE, Daniel D.; SEUNG, H. Sebastian. Algorithms for non-negative matrix factorization. En Advances in neural information processing systems. 2001. p. 556-562. [Google Scholar]

51. FITZGERALD, Derry; CRANITCH, Matt; COYLE, Eugene. Shifted non-negative matrix factorization for sound source separation. In IEEE / SP 13th Workshop on Statistical Signal Processing, 2005 . IEEE, 2005. p. 1132-1137. [Google Scholar]

52. JAISWAL, Amit. Non-Negative Matrix Factorization Based Algorithms to Cluster Frequency Basis Functions for Monaural Sound Source Separation. 2013. [Google Scholar].

53. SANCHEZ, I .; VIZCAYA, C. Tracheal and lung sounds repeatability in normal adults. Respiratory medicine , 2003, vol. 97, no 12, p. 1257-1260. [Google Scholar].

54. Breath Sounds Methodology Written by Noam Gavriely Chapter 11 breath-to-breath variability pg149

55. The R.A.L.E repository. Available on website: http://www.rale.ca/

56. Stethographics pulmonary and heart sounds to study about lung and heart sounds . Available on website : http://www.stethographics.com/

57. Lung sounds samples recorded through a 3m Littman stethoscope. Available on website :https://www.3m.com

58. ICBHI 2017 ChallengeRespiratory Sound Database. Available on website : https://bhichallenge.med.auth.gr/

59. Respiratory samples from Lippincott nursing center Available on website : https://www.nursingcenter.com

60. Thinklabs one Digital Stethoscope Available on website: https://www.thinklabs.com

61. Thinklabs Digital Stethoscope safe Auscultation respiratory samples Available on youtube: https://www.youtube.com/channel/UCzEbKuIze4AI1523_AWiK4w

62. Medscape breath sounds Available on website: https://emedicine.medscape.com/article/1894146-overview#a3

63. ERS e-learning resources
Available on website : https://dev.ers-education.org/home/

64. Samples from Respiratory Wiki
Available on website: http://respwiki.com/Breath_sounds

65. Easy Auscultation Available on website: https://www.easyauscultation.com

66. CSU Auscultation Library Available on Website: http://www.cvmbs.colostate.edu/clinsci/callan/

67. East Tennessee State University Repository
Available on website : http://faculty.etsu.edu/

68. E. Vincent, R. Gribonval and C. Févotte, Performance measurement in blind audio source separation, *IEEE Trans. Audio, Speech and Language Processing*, 14(4), pp 1462-1469, 2006.

69. C. Févotte, R. Gribonval and E. Vincent, BSS_EVAL toolbox user guide - Revision 2.0, Technical Report 1706, IRISA, April 2005.

70. SDR – HALF-BAKED OR WELL DONE? Jonathan Le Roux1 , Scott Wisdom2 , Hakan Erdogan3 , John R. Hershey2 http://www.jonathanleroux.org/pdf/LeRoux2019ICASSP05sdr.pdf

71. Taplidou, S. A., Hadjileontiadis, L. J., Kitsas, I. K., Panoulas, K. I., Penzel, T., Gross, V., & Panas, S. M. (2004, September). On applying continuous wavelet transform in wheeze analysis. In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Vol. 2, pp. 3832-3835). IEEE. [Google Scholar].

72. WEISS, Earle B.; CARLSON, C. Jeffrey. Recording of breath sounds. American Review of Respiratory Disease, 1972, vol. 105, no 5, p. 835-839. [Google Scholar].

73. FORGACS, Paul. Lung sounds. British journal of diseases of the chest, 1969, vol. 63, no 1, p. 1-12. [Google Scholar].

74. FORGACS, Paul. The functional basis of pulmonary sounds. Chest, 1978, vol. 73, no 3, p. 399-405. [Google Scholar].

75. Murphy, R. L., Vyshedskiy, A., Power-Charnitsky, V. A., Bana, D. S., Marinelli, P. M., Wong-Tse, A., & Paciej, R. (2004). Automated lung sound analysis in patients with pneumonia. Respiratory care, 49(12), 1490-1497. [Google Scholar].

76. Haider, Nishi Shahnaj, and Justin Joseph. "An investigation on the statistical significance of spectral signatures of lung sounds." (2017). [Google Scholar].

77. Hafke-Dys, H., Bręborowicz, A., Kleka, P., Kociński, J., & Biniakowski, A. (2019). The accuracy of lung auscultation in the practice of physicians and medical students. PloS one , 14 (8), e0220606. [Google Scholar]

78. Sovijarvi AR, Dalmasso F, Vanderschoot J, Malmberg LP, Righini G, Stoneman SA. Definition of terms for applications of respiratory sounds. Eur. Respir. Rev. 2000;10:597–610. [Google Scholar].

79. Sarkar, M., Madabhavi, I., Niranjan, N., & Dogra, M. (2015). Auscultation of the respiratory system. Annals of thoracic medicine, 10 (3), 158. [Google Scholar]

80. Chasin M. Musicians and the prevention of hearing loss. San Diego: Singular Publishing Group; 1996. [Google Scholar].

81. Ferns, Terry, and Susan West. "The art of auscultation: evaluating a patient's respiratory pathology." *British Journal of Nursing* 17.12 (2008): 772-777. [Google Scholar].

82. A basic review on pulmonary auscultation. Available from : https://www.rn.com/nursing-news/basic-review-of-pulmonary-auscultation/.

83. Gottlieb, Eric R., Jason M. Aliotta, and Dominick Tammaro. "Comparison of analogue and electronic stethoscopes for pulmonary auscultation by internal medicine residents." *Postgraduate medical journal* 94.1118 (2018): 700-703. [Google Scholar].

84. Grenier, Marie-Claude, et al. "Clinical comparison of acoustic and electronic stethoscopes and design of a new electronic stethoscope." *American Journal of Cardiology* 81.5 (1998): 653-656.[Google Scholar].

85. Reichert, Sandra, et al. "Analysis of respiratory sounds: state of the art." *Clinical medicine. Circulatory, respiratory and pulmonary medicine* 2 (2008): CCRPM-S530. [Google Scholar].

86. Information about wheezes. Available on Wikipedia.

87. Sengupta, Nandini, Md Sahidullah, and Goutam Saha. "Lung sound classification using cepstral-based statistical features." *Computers in biology and medicine* 75 (2016): 118-129.[Google Scholar].

88. R. Palaniappan, K. Sundaraj, N. U. Ahamed, A. Arjunan, S. Sundaraj,Computer-based respiratory sound analysis: a systematic review, IETETechnical Review 30 (3) (2013) 248–256. [Google Scholar]

89. P. Forgacs, The functional basis of pulmonary sounds., Chest73 (3) (1978)399–405. [Google Scholar]

90. P. Piirila, A. Sovijarvi, Crackles: recording, analysis and clinical significance, European Respiratory Journal 8 (12) (1995) 2139–2148. [Google Scholar].

91. Forgacs, Paul. "Crackles and wheezes." *The Lancet* 290.7508 (1967): 203-205 [Google Scholar].

92. Andres, Emmanuel. "Respiratory Sounds Analysis in the World of Health 2.0 and Medicine 2.0." *EC Pulmonology and Respiratory Medicine* 7 (2018): 564-585. [Google Scholar].

93. Upadhyay, Navneet, and Abhijit Karmakar. "Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study." *Procedia Computer Science* 54 (2015): 574-584. [Google Scholar]

94. D. FitzGerald, B. Lawlor, and E. Coyle, "Prior subspace analysis for drum transcription," in Proceedings of the Audio Engineering Society Convention, Amsterdam, The Netherlands, 2003.

95. Chen, Hai, et al. "Automatic Multi-Level In-Exhale Segmentation and Enhanced Generalized S-Transform for wheezing detection." *Computer methods and programs in biomedicine* 178 (2019): 163-173. [Google Scholar]

96. Harmonic and Percussive Source Separation Available at : https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/02-teaching/2017w_mpa/LabCourse_HPSS.pdf.

97. Canadas-Quesada, F., Ruiz-Reyes, N., Carabias-Orti, J., Vera-Candeas, P., and Fuertes-Garcia, J. A non-negative matrix

factorization approach based on spectro-temporal clustering to extract heart sounds, Applied Acoustics, vol. 125, pp. 7-19, 2017. [Google Scholar]

98. Canadas-Quesada, Francisco Jesus, et al. "Percussive/harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints." *EURASIP Journal on Audio, Speech, and Music Processing* 2014.1 (2014): 26. [Google Scholar]

99. Canadas-Quesada, Francisco J., et al. "Harmonic-percussive sound separation using rhythmic information from non-negative matrix factorization in single-channel music recordings." *Proc. Intl. Conf. Digital Audio Effects (DAFx), Edinburgh, UK* . 2017. [Google Scholar]

100. Essid, Slim. "Matrix Co-Factorisation and Applications to Music Analysis." *Machine Learning for Music Discovery Workshop, International Conference on Machine Learning (ICML) 2017*. 2017.[Google Scholar]

101. Seichepine, Nicolas, et al. "Soft nonnegative matrix co-factorization." *IEEE Transactions on Signal Processing* 62.22 (2014): 5940-5949. [Google Scholar]

102. De La Torre Cruz, J., Cañadas Quesada, F. J., Ruiz Reyes, N., Vera Candeas, P., & Carabias Orti, J. J. (2020). Wheezing Sound Separation Based on Informed Inter-Segment Non-Negative Matrix Partial Co-Factorization. *Sensors*, *20*(9), 2679.[Google Scholar]

9 APPENDEX

9.1 Development of user-friendly interface

Variable Nomenclature

KR : Representation for number of common respiratory basis vectors (K_R) that has to be entered by user in GUI model

KW: Representation for number of individual wheezing basis vectors (K_W) that has to be entered by user in GUI model

Lambda : Representation for (λ) value that has to be entered by user in GUI model

X : Name of the magnitude spectrogram plot of respiratory mixture signal X in GUI model

XR : Name of the magnitude spectrogram representation of estimated respiratory signal $\hat{X}_R$ in GUI model

XW : Name of the magnitude spectrogram representation of estimated wheezing signal $\hat{X}_W$ in GUI model.

As already mentioned in Section 3 one of the objectives of our thesis study is to create a GUI model. The GUIs (also known as graphical user interfaces or UIs) provide point-and-click control of software applications, making easy for a standard user to use more complex instructions and learn the specific language in order to run the application. The main aim of developing GUI is to provide the easiest possible way for physicians to directly interact with the computerized respiratory sounds. Through the GUI interface physicians can upload the recorded respiratory sound and run the GUI interface by allocating different parameters into the parameters space provided. By running the GUI interface the proposed ST-NMPCF algorithm written as a code gets executed and the physicians will be able to visualise the spectrograms of separated respiratory sound and separated wheezing sound from the mixture signal provided.

The spectro-temporal visualization of the respiratory sounds is very important for physicians to accurately recognise the patient's respiratory disorders irrespective of their age, gender and demography. The reason

behind representing the respiratory sounds in their frequency-time domain representation is they show acoustic symptomatic characteristics, indicate the presence of abnormalities in the respiratory system along with severity and the location of the most frequently found airway obstructions in asthma and respiratory stenoises. From the spectrogram representation of wheeze and respiratory signal the physician can analyse the data more accurately and can proceed with the diagnosis of the patient at an early stage.

The screen capture of our developed GUI interface is depicted in ***Figure 9.1.1***. The GUI interface is provided with few parameters for the user to enter in order to view the results. Each parameter is allocated with its respective text boxes which allows the user to enter the optimal values from ***Table 5.5.2.***

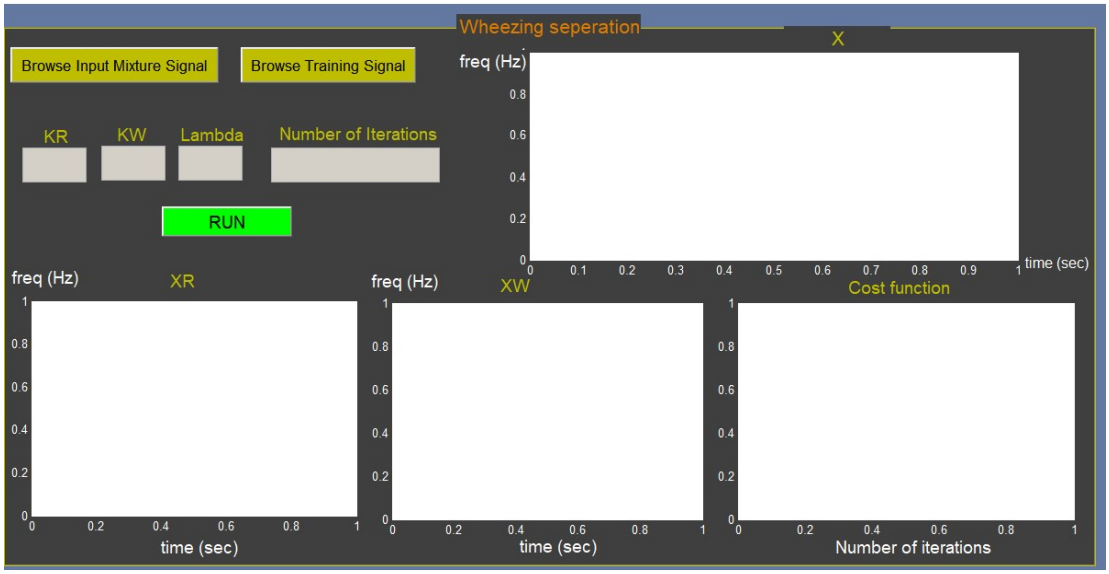


Figure 9.1.1 Screen capture of the created Graphical User Interface

In this section we are going to discuss the working of the GUI interface in four steps:

1. Access to required signals
2. Allocation of parameter values
3. Execution
4. Viewing the results

1. Access to required signals

From the explanation given in section 4.3 that explains about ST-NMPCF approach it is clear that for the proposed algorithm to run we need two signals. The two signals are input mixture signal $x[n]$ and the respiratory training signal $y[n]$. To get access to these signals, two push buttons are provided in the created GUI interface model, named Browse Input Mixture Signal and Browse Training Signal on top left in green color that can be observed in ***Figure 9.1.1.*** The two push buttons allows the user to select the signal $x[n]$ and $y[n]$.

To select Input mixture signal $x[n]$:

To get access to the input mixture signal $x[n]$ the user has to press the push button named Browse Input Mixture Signal. On pressing the push button a "File Selector" dialogue box opens as shown in ***Figure 9.2***

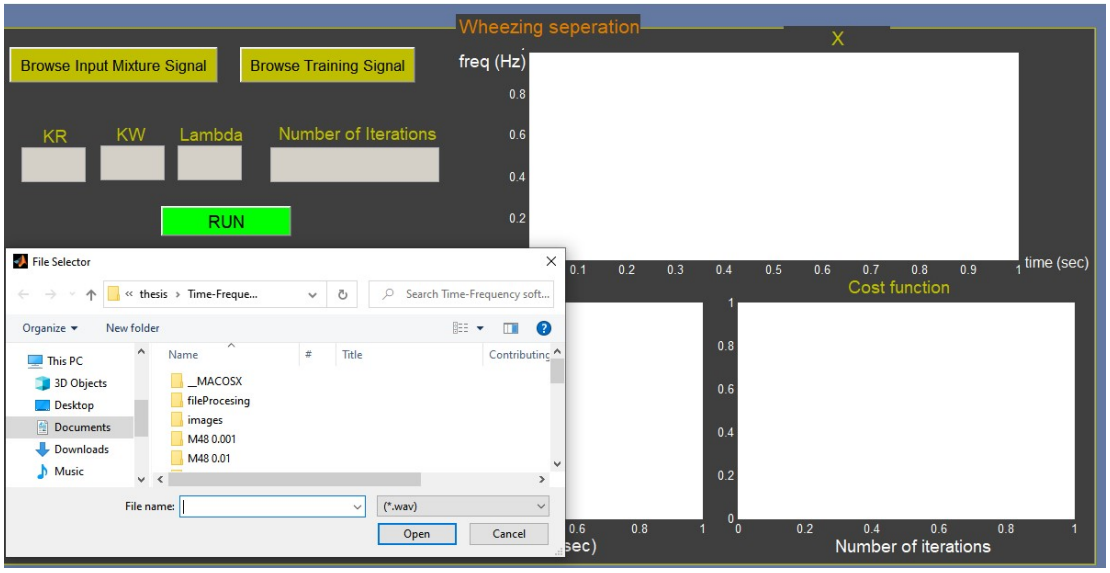


Figure 9.1.2 File selctor dialogue box on pressing the Browse Input Mixture Signal push button.

Now the user has to select the input respiratory mixture signal $x[n]$ from which he wants to separate the wheezing sounds and press open to load the file.

To select the training signal $y[n]$:

Next step is to have access to the training signal or prior knowledge signal $y[n]$**.** As our proposed ST-NMPCF approach is a supervised approach it uses training signal to separate the target sources from the input mixture. Therefore in our GUI model we created a browsing button named Browse Training Signal to select training signal $y[n]$ to proceed

with the separation process. The training signal has to be composed of solo-respiratory sounds concatenated end-to-end. By pressing the push button Browse Training Signal a File Selector dialogue box similar to as shown in ***Figure 9.1.2*** opens. The user has to select the required signal and press open to load the training signal $y[n]$ in to ST-NMPCF algorithm.

2. Allocation of parameter values

As discussed in ***Table 5.5.1*** to run the ST-NMPCF algorithm we need several parameters whose values must be assigned by the user to get the separation results. The parameters that the user needs to enter are K_R, K_W, λ_1 and number of iterations. In our GUI model we provided parameter space for all the parameters where the user can enter the values for the parameters. Explanation about each one is given below:

K_R :

K_R defines about the number of respiratory components that have to be assigned by the user to collect all the respiratory sounds from the input mixture signal selected in step 1. Based on this number the proposed algorithm evaluates all the temporal and spectral respiratory basis vectors and collects them in common respiratory basis matrix A_R. In order to assign this value in the GUI model, we created a text box named with KR (on top left indicated by a red colour square box ***Figure 9.1.3***)

Figure 9.1.3 GUI model with markings on different parameters

K_W :

K_W defines about the number of wheezing components that have to be assigned by the user to collect all the wheezing sounds from the input mixture signal selected in step 1. Based on K_W number the proposed algorithm evaluates all the spectro-temporal characteristics of wheezing individual basis vectors and collects them in harmonic wheezing basis matrix A_W. In order to assign this value in the GUI model, we created a text box named with KW (on top left indicated by a blue colour square box ***Figure 9.1.3***)

λ_1 (lambda):

As mentioned in section 5.3 in the ***Table 5.5.1*** we need lambda value to control the reconstruction error of each input matrix's segmented column-blocks and respiratory training signals. Therefore in our GUI model we allocated a entry box to allow the user to enter the λ_1 value that belongs to one of these values in the array [0.001, 0.01, 0.1, 1, 10]. The text box to assign $\lambda_1$value in interface is indicated in orange colour square box named as Lambda in ***Figure 9.1.3***. For almost all signals that we analysed during the optimisation process and testing process the value of 0.001 gave its best results. But the user can visualize the difference in separation performance by varying each value.

Number of iterations:

For obtaining good separation results the objective function equation (4.3.3.1) has to converge to its best minimum value. The convergence of objective function is obtained by iterating the factor matrices A_R, A_W, S_W, S_R with several number of iterations by using the update rules derived in equations (4.3.3.10 to 4.3.3.13). To allocate the value for number of iterations using GUI interface we created a text box in the GUI model indicated in a enclosed pink colour box in ***Figure 9.1.3*** named as Number of Iterations. The user can view separation performance by varying the number of iterations by entering the value in the allocated text box.

3. Execution

In order to view the acquired results the program has to be executed. To proceed with the execution process the user has to follow step 1 and step 2 as explained above and enter all the parameter values of KR, KW, Lambda and Number of Iterations as shown in ***Figure 9.4.***

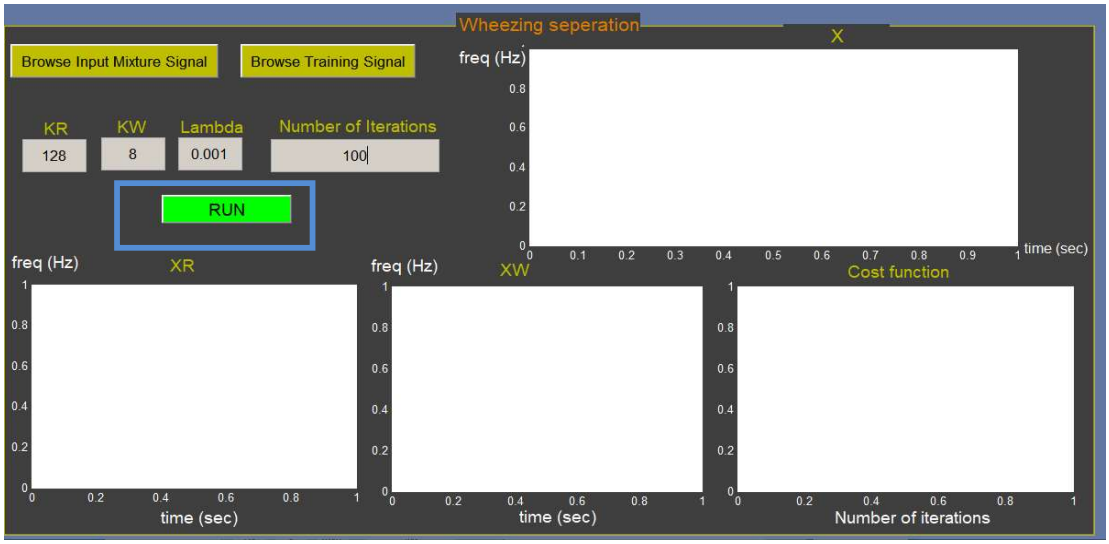


Figure 9.1.4 Screen capture of GUI model with allocated parameter values

The user has to push the green coloured RUN button which is indicated by blue coloured box in ***Figure 9.1.4*** to execute the program and view the results.

4. Viewing the results

After executing the MATLAB program through created GUI model, the user has to examine the results through spectrogram representation of estimated wheezing and estimated respiratory signals. In order to present different plots on GUI model MATLAB provides axes tool which on programmed shows the desired results that the programmer want to show. So by using this axes tool in MATLAB we created four axes plots named as X, XR, XW and Cost function. Each axes plot has its own importance and shows different spectrogram representations which helps the user in easy examination. To explain about importance of each axes plot to the user we indicated axes plots with numbers 1, 2, 3 and 4 as shown in ***Figure 9.1.5***

From ***Figure 9.1.5*** axes plot labelled with 1 is allocated for representing the spectrogram of input mixture signal x[n] selected in step 1 i.e. the axes plot is used for representing the time-frequency characteristics of respiratory mixture signal. We named the axes plot as X in GUI model to make the user identify the mixture signal

Figure 9.1.5 Indication of axes plots in GUI model with numbers

From ***Figure 9.1.5*** axes plot labelled with 2 is allocated for representing the spectrogram of estimated respiratory signal $\hat{X}_R$ obtained by running the ST-NMPCF algorithm i.e. the axes plot is used for representing the time-frequency characteristics of estimated respiratory signal. We named the axes plot as XR in GUI model to make the user identify the estimated respiratory signal.

From ***Figure 9.1.5*** axes plot labelled with 3 is allocated for representing the spectrogram of estimated wheezing signal $\hat{X}_W$ obtained by running the ST-NMPCF algorithm i.e. this particular axes plot is used for representing the time-frequency characteristics of estimated wheezing signal. We named the axes plot as XW in our GUI model to make the user identify the spectrogram of the estimated wheezing signal. As all the axes plots labelled 1,2 and 3 are time-frequency plots, we labelled the y-axes as frequency (Hz) and x-axes as time (seconds)

The axes plot labelled 4 in ***Figure 9.1.6*** is allocated for representing the cost function of the ST-NMPCF approach i.e. the plot tells the user, at which iteration the values of the all the matrices A_R, A_W, S_R and S_W are converging to their optimal value or getting saturated to one particular value. Therefore the user can stop the iteration process at that particular iteration value while running the algorithm rather than running for 100 iterations. We named the axes plot as Cost function in our GUI model to make the user identify the correct plot from the interface.

In order to demonstrate how our created GUI model portray the results we followed steps 1 to 3 as mentioned above. At first we selected the input mixture signal $x[n]$ and respiratory training signal $y[n]$ and then allocated the optimal parameter values for $K_R = 128$, $K_W = 8$, and $\lambda_1 = 0.001$ obtained from ***Table 5.5.1*** within the allocated spaces of KR, KW, Lambda and Number of Iterations in the GUI model. The results are depicted in ***Figure 9.1.6***

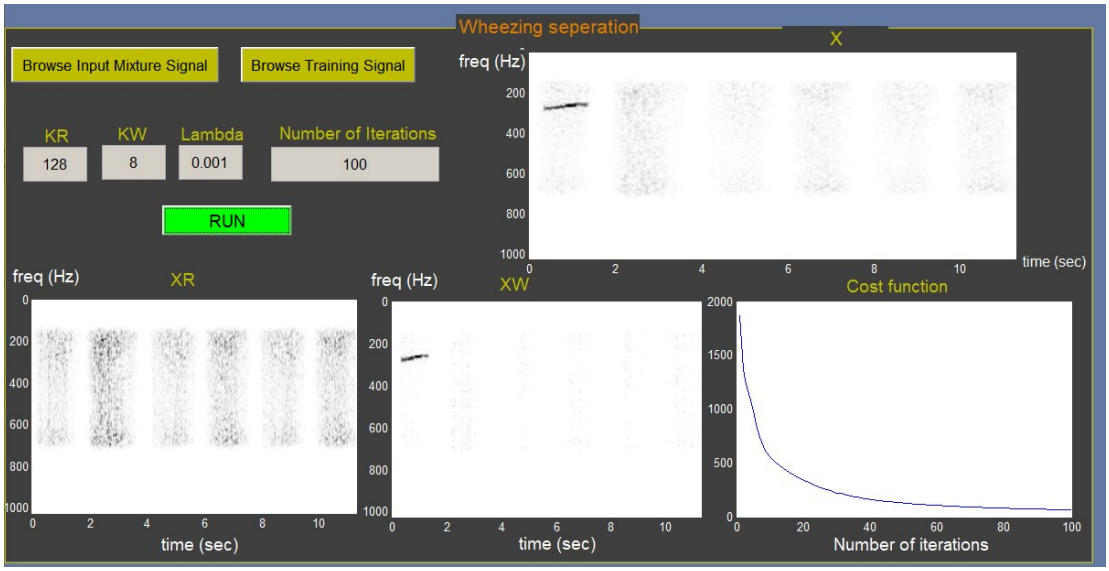


Figure 9.1.6 Screen capture of separation results of a mixture signal obtained on GUI model

From ***Figure 9.1.6*** we can observe that the axes plot named with X represents the spectrogram of the selected input respiratory mixture signal which is composed of both respiratory and wheezing sounds. The axes plot named with XR represents the spectrogram of estiamted respiratory signal without any wheezing sound in it. Similarly the plot named with XW represents the spectrogram of estimated wheezing signal. Finally the axes plot named as cost function shows the convergence of the algorithm. Therefore from the cost function plot we can conclude that the algorithm is converging form iteration 40 which implies by allocating 40 iterations or more we can obtain the estimated signals.