WHEEZE DETECTION IN RESPIRATORY SOUNDS VIA STATISTICAL SIGNAL MODELING

by

Sergül Aydöre

B. S. Electrical and Electronics Engineering, Boğaziçi University, 2007

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical and Electronics Engineering
Boğaziçi University

2009

WHEEZE DETECTION IN RESPIRATORY SOUNDS VIA STATISTICAL SIGNAL MODELING

APPROVED BY:

Assoc. Prof. M. Kıvanç Mıhçak (Thesis Supervisor)	.....
Prof. Yasemin P. Kahya (Thesis Co-supervisor)	.....
Assoc. Prof. Ata Akın	.....
Asst. Prof. Murat Saraçlar	.....
Asst. Prof. Serdar Kozat	.....

DATE OF APPROVAL: 15.06.2009

ACKNOWLEDGEMENTS

I, firstly, thank my advisor Assoc. Prof. M. Kıvanç Mıhçak for teaching me how to conduct research. I appreciate his insightful approach to the problems, invaluable help, patience and encouragement.

I also thank my co-advisor Prof. Yasemin P. Kahya for her significant help during the work, intimate attitude and understanding.

Secondly, thanks to my friend İpek Şen from our laboratory for supporting not only during my studies but also during difficulties in my life. Her experience and insights in this area played a key role when solving the problems I was faced with.

I am grateful to my thesis committee members Assoc. Prof. Ata Akın, Asst. Prof. Murat Saraçlar and Asst. Prof. Serdar Kozat for their outside help, support and broadening my horizons.

I owe special thanks to my dear friends Filiz Ateş, Nargiz Kalantarova, Sevinç Mutlu, Nermin Topaloğlu, Didar Talat, Helin Dutağacı, Berçem Dutağacı, Dan Gareau, Ercan Atam and Pouyan Mohajerani for their close friendship and sharing my troubles during the preparation of this thesis. I also thank to all members of BUSIM for their practical help and guidance.

I would like to thank TUBITAK for supporting the financial means of my graduate study.

My deepest gratitude goes to my family for their endless love at the all stages of my life.

ABSTRACT

WHEEZE DETECTION IN RESPIRATORY SOUNDS VIA STATISTICAL SIGNAL MODELING

The aim of this study is detection of wheeze and non-wheeze epochs within respiratory sound signals acquired from patients with asthma and COPD. Since a wheeze signal, having a sinusoidal waveform, has a different behavior in time-frequency domain from that of a non-wheeze signal, the features selected for detection are Renyi entropy, f_{50}/f_{90} ratio and mean-crossing irregularity. Upon calculation of these features for each wheeze and non-wheeze portion, two approaches are proposed. In the first approach (multi-dimensional approach) the whole data scattered as two classes in three dimensional feature space are assumed to be Gaussian distributed. Then, a decision rule is applied for two different Gaussian random vectors. In the second approach (one-dimensional approach) the three dimensional data are projected onto the single dimensional space that separates the two classes best by using Fisher Discriminant Analysis (FDA). These one-dimensional data are also assumed to be Gaussian distributed. Then, a decision rule is applied for two different Gaussian variables. Finally when the total number of false positives and false negatives are considered, the minimum value of probability of error over the data set is found to be 0.045 for both approaches and the Chernoff upper bounds for this error are found as 0.09 and 0.10 for multi and one-dimensional approaches, respectively.

ÖZET

SOLUNUM SİNYALLERİDEKİ ÜFÜRÜMÜN İSTATİSTİKSEL SİNYAL MODELLEME İLE SEZİMİ

Bu çalışmanın amacı astım ve KOAH hastalarından alınmış olan solunum sesi sinyallerindeki üfürümlü ve üfürümsüz bölgelerin sezimidir. Üfürümlü sinyaller sinüs benzeri özellik gösterdiklerinden dolayı, zaman-sıklık uzayında üfürümsüz sinyallerden farklı özellikler gösterirler. Bu bağlamda, sezim için Renyi entropisi, f_{50}/f_{90} oranı ve ortalama kesme düzensizliği öznitelikleri tanımlanmıştır. Her üfürümlü ve üfürümsüz bölgeler için hesaplanmış olan öznitelikler üzerinden iki yöntem önerilmiştir. İlk yöntemde (çok boyutlu yöntem), üç boyutlu uzayda dağılmış olan üfürümlü ve üfürümsüz bölgelerin normal dağılım sergiledikleri kabul edilmiştir. Bu iki farklı normal rasgele vektörler için karar kuralı önerilmiştir. İkinci yöntemde (tek boyutlu yöntem), üç boyutlu veri Fisher Çözümleme tekniği ile iki sınıfın en iyi ayrıldığı tek boyuta indirgenmiş ve bu tek boyutlu verinin de normal dağılım gösterdiği kabul edilmiştir. Daha sonra bu iki farklı normal rasgele degişkenleri için bir karar kuralı önerilmiştir. Sonuç olarak, yanlış red ve yanlış kabul değerlerinin toplamı dikkate alındığında, iki yöntem için de elimizdeki veriler üzerinden hesaplanan en küçük hata olasılığı 0.045 olarak bulunmuştur. Bu değerin üst sınırı çok boyutlu yöntem için ve tek boyutlu yöntem için sırasıyla 0.09 ve 0.10 olarak bulunmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

j	Class number, 1/0 for wheeze/non-wheeze class
$\bar{K}$	Kurtosis value
N	Total number of windows from each class
P_e	Probability of error
P_F	Probability of false alarm
P_M	Probability of miss
$\mathbf{w}$	Fisher Discrimination vector
$y_{j,i}$	Projected value of i^{th} sample from class j
Y_j	One-dimensional random variable of projected values for class j
$\mathbf{x}_j$	Specific realization of the random vector $\mathbf{X}_j$
$\mathbf{X}_j$	Three-dimensional random vector of class j
$\bar{\mu}_j$	Three-dimensional mean vector of random vector $\mathbf{X}_j$
μ_j	One-dimensional mean value of random variable Y_j
$\mathbf{\Sigma}_j$	Covariance matrix with size 3×3 of random vector $\mathbf{X}_j$
σ_j^2	Variance of random variable Y_j
$\mathcal{X}_j$	The alphabet of random vector $\mathbf{X}_j$
$\mathcal{Y}_j$	The alphabet of random variable Y_j

ATS	American Thoracic Society
COPD	Chronic Obstructive Pulmonary Disease
CORSA	Computerized Respiratory Sound Analysis
DAQ	Data Acquisition
FDA	Fisher Discriminant Analysis
FFT	Fast Fourier Transform
KL	Kullback Leibler
ML	Maximum Likelihood
p.m.f	Probability Mass Function
pdf	Probability Density function

psd	Power Spectral Density
ROC	Receiver Operation Characteristic

1. INTRODUCTION

Listening to pulmonary sounds for diagnosis date backs to ancient times. There have been some writings remained from Hippocratic School (B. C. 400) claiming that some abnormal sounds were heard upon the chest of the patients [1]. The idea of using the stethoscope to hear the pulmonary sounds upon the chest came up and the most primitive stethoscope was a narrow tube created by a piece of paper [1]. Since then, there have been many improvements on the stethoscope.

Today, two main methods used in the diagnosis of respiratory diseases are computerized imaging techniques and auscultation, the latter being inexpensive, noninvasive, simple and patient-friendly. In many respiratory disorders (e.g. pneumonia, emphysema, bronchiectasis, COPD and asthma), adventitious sounds specific to the respiratory disorder are observed in the sound data, and this generally provides invaluable information that can eliminate the need for imaging techniques. However, stethoscope permits the signals below 112 Hz and attenuates the rest whereas the human ear is not much sensitive to the frequencies below this value [2]. The upper frequency is 2000 Hz and 6000 Hz for normal respiratory sound signals and adventitious sounds, respectively [3]. Hence, the stethoscope does not perform well within the frequency band where the human ear is most sensitive and the respiratory sounds are located. Moreover, this method highly depends on the subjective evaluation of the physician and it is impossible to make an objective quantification. Lastly, since the sounds can not be recorded, it is impossible to represent the respiratory sounds graphically by stethoscope.

In order to make auscultation a more valuable diagnostic tool, computerized methods have been applied more so in recent studies. The first studies in this area are summarized in [2] and [4]. The computerized auscultation and analysis methods of respiratory sound signals can be summarized as capturing the respiratory sounds as analog signals via a transducer (or an array of transducers), digitization after preprocessing, and interpretation of the digitized data using various analysis techniques

in the computer environment so the sound data can be represented graphically, data banks can be constructed with minimal frequency content lost. Therefore, the data acquired with computerized methods are more reliable, useful and non-dependent on the subjectivity of the physician.

The pulmonary sounds are produced by the turbulence of the air in the airways of the lungs. Normal breath sound is the sound which can be heard from the chest wall of the healthy subjects. It is acoustically soft and may contain many frequency components [1], [2], [5]. However, its spectrum may change and some additional abnormal sounds would be heard if there are changes in the lung structure which will occur in pathological cases [2], [6], [7]. Wheezes and crackles are the abnormal sounds the analysis of which may be helpful in the diagnosis of respiratory diseases [4]. Wheezes are continuous adventitious lung sounds which may be occurred in patients with obstructive airway diseases, especially in acute episodes of asthma patients. Crackles are discontinuous abnormal sound which may be auscultated in patients with cardiorespiratory diseases.

Wheeze signals (having a continuous and sinusoidal waveform) which are clinical indicators of obstructive respiratory diseases (e.g. COPD, asthma) are aimed to be detected in this study. When the combination of intrathoracic pressure, airway narrowing and air velocity reaches a critical degree, wheezing will occur [1]. The mechanism of wheeze production is given in Figure 1.1.

The existence of a wheeze and its characteristics (e.g. main frequency component, or ratio of its time duration to the total breath cycle) can be accepted as an indicator of the degree of the bronchial obstruction [5]. There are several suggestions on frequency component and time duration of the wheeze. According to ATS wheeze has a main frequency of 400 Hz or more and its' time duration is not more than 250 ms [5]. On the other hand, CORSA suggests that the main frequency and time duration of wheeze are larger than 100 Hz and 100 ms, respectively [7]. Some researchers claim that the time duration of the wheeze can be as short as 30 ms [8]. Thus, the frequency range of the wheeze is from 80 Hz to 2000-2500 Hz according to Gavriely [1] and from 100 Hz

to 1000 Hz according to Pasterkamp [2].

In this study, two different methods have been proposed for the detection of a wheeze signal. The methods are intended to be a solution to the detection problem of a wheeze, i.e. whether it exists or not given a segment of recorded lung sound signal. The existence of a wheeze and its characteristics (e.g. main frequency component, or ratio of its time duration to the total breath cycle) can be accepted as an indicator of the degree of the bronchial obstruction [5]. This relationship should be especially emphasized since it reveals the importance of the detection problem. Before the development of classification algorithms, three different features which are expected to be distinctive are extracted from the signal windows which after a visual inspection by an expert have been labeled as wheeze and non-wheeze on the respiratory sounds taken from asthma patients. Finally, proposed methods are applied in this feature space.

The features proposed are Renyi entropy and mean-crossing irregularity calculated in the time domain, and f_{50}/f_{90} ratio calculated in the frequency domain. After feature extraction, the wheeze and non-wheeze portions can be considered each as elements of one of the two vector sets in the three-dimensional feature space.

First method is the multi-dimensional approach which is to fit a Gaussian distribution over the three-dimensional feature space and apply a decision rule. For the second approach which is the one-dimensional approach, this three-dimensional space is reduced to one-dimensional space by FDA. Then, a decision rule is selected after Gaussian fitting in this one-dimensional space. The Gaussianity assumption of the multi and one-dimensional data is verified utilizing tools from information theory. Then, performance rates of the methods are evaluated in theoretical and experimental sense. Over the 245 wheeze and 245 non-wheeze windows for each, the performance rates of the approaches are found as %95.5 both for multi and one-dimensional data. Our results confirm the validity and efficiency of the proposed methods and features.

In similar studies carried previously by other researchers, artificial neural networks were generally adopted; and either the original signal itself and its Fourier trans-

form coefficients [9], or power spectrum components [10-12], or, descriptive statistical features derived from wavelet transform coefficients [10] have been employed as features. Moreover, there are some studies which have applied peak detection on time-frequency representations like spectrogram or continuous wavelet transform to identify the existence of a wheeze [13-16]. Our study utilizes the statistical approach by merging the concepts from detection and estimation theory, pattern recognition and information theory which distinguishes its approach from previous works.

The thesis is organized as follows: In chapter 2 the data acquisition system and the source of the data are given. Three proposed features are explained in chapter 3. In chapter 4, multi dimensional approach is explained theoretically, Gaussianity of multi-dimensional results is verified and results of the approach are given. One-dimensional approach is presented along with dimensionality reduction, verification of Gaussianity of one-dimensional data and results of the approach in chapter 5. Then, upper bounds of the probability of errors for both approaches are given in chapter 6. In chapter 7, the results of the two approaches are discussed and compared. Cocnlusions are drawn in chapter 8.

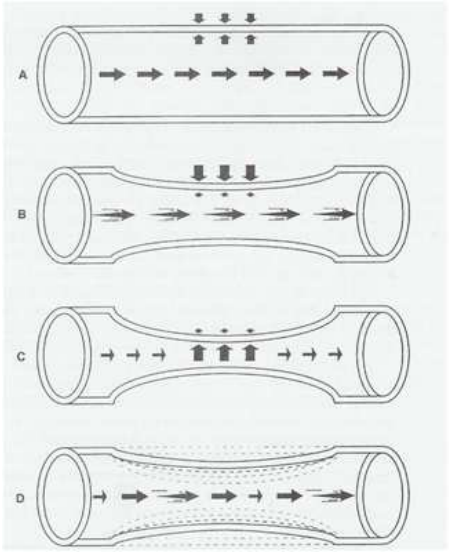


Figure 1.1. Mechanism of wheeze production [17]

2. EXPERIMENTAL SETUP AND DATA

The data in this study is obtained from the data base which has been formed by the lung sound analysis system [18] with 14 channels. This system has been developed in the Boğaziçi University Lung Sound Laboratory. The system is composed of 14 air-coupled electret microphones (SONY ECM-44 BPT) placed on the posterior chest wall. Analog amplifier-filter unit with gain 100 and 8th order Butterworth low-pass filter with 4kHz cut-off frequency and 6th order Bessel high-pass filter with 80 Hz cut-off frequency are used to minimize frictional noise and heart sound interference with minimal phase distortion, i.e. the unit regularizes the signal at the output of the microphone. A 12-bit ADC card (National Instruments DAQCard-6024E) digitalizes the regularized signal at a 9.6 kHz sampling rate and transmits it to a notebook computer which works with an interface identified in the LabView media. Fleisch type pneumatachograph (Validyne CD379) is used in order to synchronize signals with the flow cycle. The duration of one data acquisition session is 15 seconds and data from a single patient are taken in more than one sessions.

The data in this study was taken from asthma and COPD patients under treatment in Istanbul Yedikule Teaching Hospital for Chest Diseases and Thoracic Surgery with the guidance of a physician specialized in pulmonary medicine. Moreover, informed concent was taken from the subjects before experiments. There were 4 men and 3 women subjects in the age of 50 ± 17. For each subject, the data sessions containing at least one wheeze during 15 seconds are selected. Equal number of wheeze and non-wheeze windows are labeled from the selected data. 245 wheeze and 245 non-wheeze signals are used in this study. The experimental setup and DAQ system are depicted in Figure 2.1-a and b.

(a)

(b)

Figure 2.1. Experimental Setup (a) The components of the system (b) Data acquisition from the subject

3. FEATURE EXTRACTION

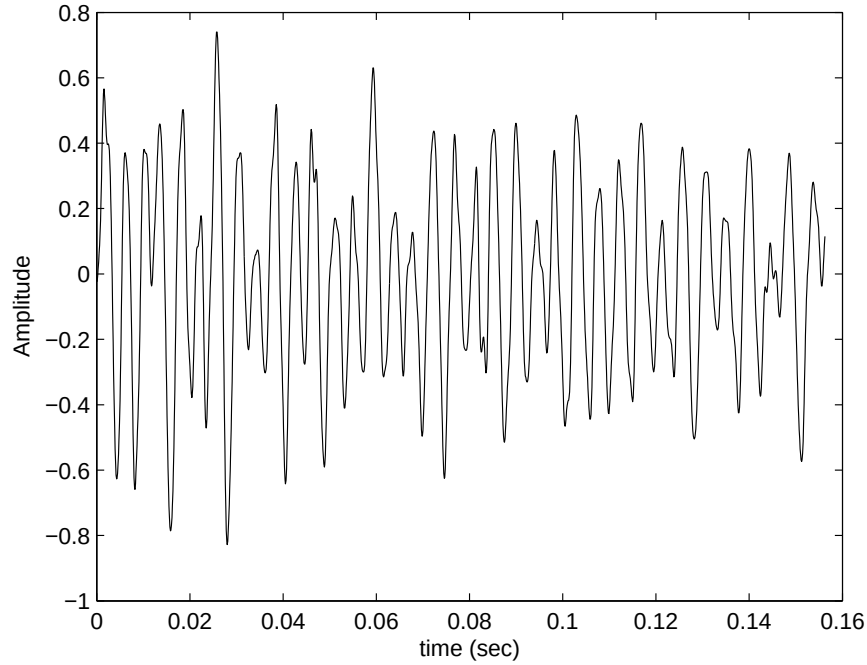
The wheeze and non-wheeze signals display distinct properties both in time and frequency domains. Wheeze signals have more sinusoidal characteristics compared to non-wheeze signals. Samples of each type of signal in time domain are given in Figure 3.1. Inspiring from these different characteristics of the signals in time-frequency domain, three different features are defined in order to classify the two classes of the signals. Those features are calculated for each signal window which is labeled as wheeze or non-wheeze. Then, two classes of the signals can be considered as two sets in three-dimensional space with the number of elements 245 for each. For these three-dimensional sets, two approaches are proposed to detect the signals. One of the approach is to fit three-dimensional Gaussian pdf for each class and then set up a decision rule from those pdf's. The other approach is to reduce the dimension of three to one dimension by FDA and then apply a decision rule for the pdf's of one-dimensional data which are assumed to be Gaussian. Finally those two approaches are compared and evaluated in terms of the probability of errors and the number of false positives and false negatives.

Note that, in the following sections, E(.), Var(.), log(.) D(.||.), $\mathbf{1}_{\{.\}}$ and $< \mathbf{a}, \mathbf{b} >$ are used to denote expectation of a given random variable, variance of a given random variable, natural logarithm, KL divergence, standard indicator function and the dot product between two vectors, respectively, and for a given matrix $\mathbf{A}$, $\mathbf{A}^T$ represents its transpose.

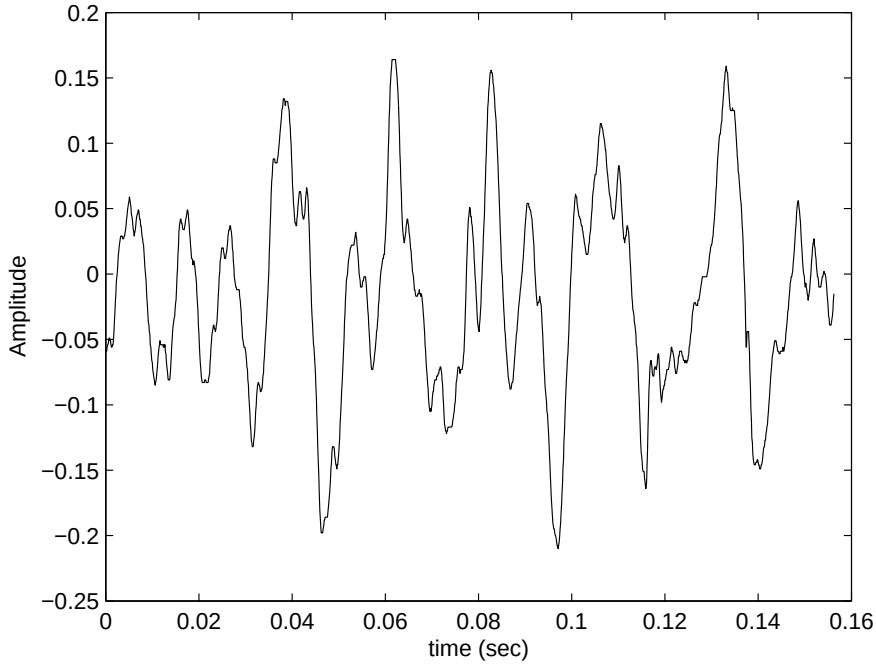
3.1. Renyi Entropy

Renyi Entropy is a measure of an uncertainty, diversity and randomness of a system. It is defined as:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} \mathrm{p}_i^\alpha \right) \tag{3.1}$$



(a)



(b)

Figure 3.1. Data in time domain (a) Wheeze (b) Non-wheeze

of order α for $0 < \alpha < \infty$, $\alpha \neq 1$ for a discrete random variable X. Shannon entropy function can be obtained if the limit $\alpha \longrightarrow 1$ is taken. The distribution of the values of wheeze and non-wheeze signals in time domain are considered as p.m.f by

calculation and then normalization of the histograms. Then, Renyi entropy of order two is computed over those p.m.f's. Since the wheeze signals have sinusoidal character, their p.m.f's should be more like uniform distribution. On the other hand, non-wheeze signals have more irregular character so their p.m.f's should be closer to the normal distribution. Thus, Renyi entropy values for wheeze signals are expected to be larger than that of non-wheeze signals. The histograms of the values of Renyi Entropy of order two with the bin number of 16 of both classes are shown in Figure 3.2. According to the Figure, it can be seen that the results are consistent with the expectation that wheeze windows have larger Renyi entropy values than non-wheeze signals.

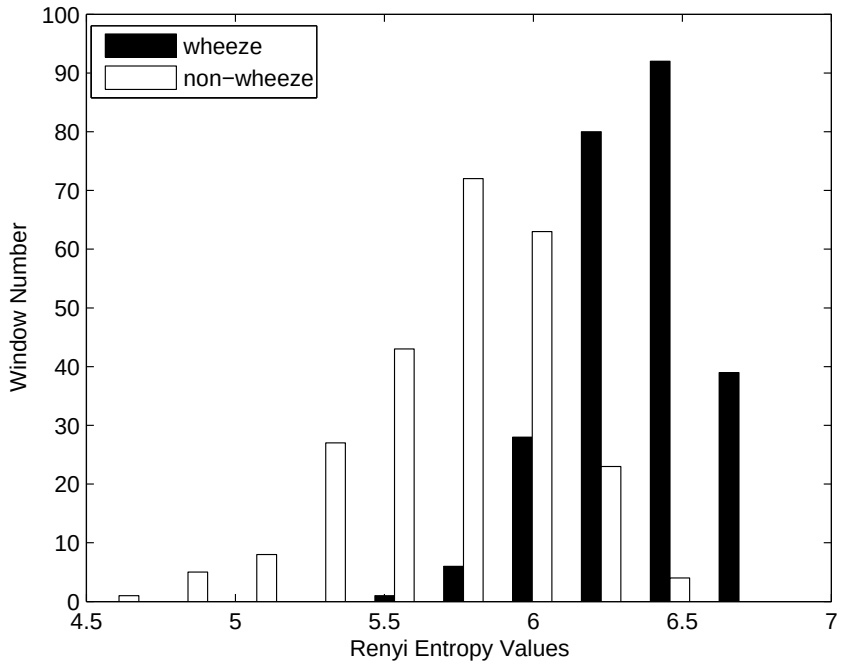
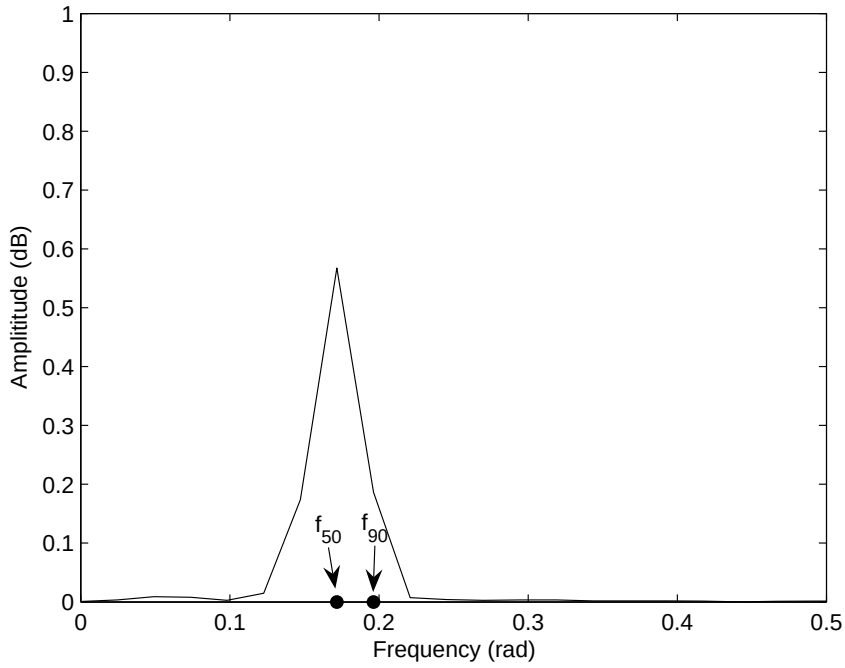


Figure 3.2. Histograms for Renyi Entropies

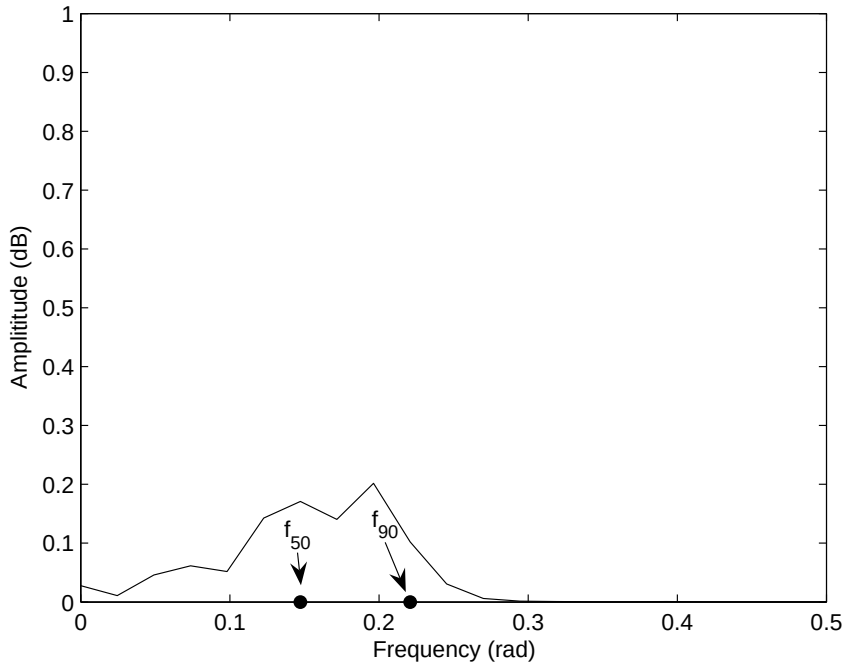
3.2. f_{50}/f_{90} Ratio

f_{50} denotes the frequency where the ratio of the area under the psd to the total area is 50 % and f_{90} is the frequency where the ratio is 90 %. When the psd is estimated very close to the real value, since the power density of wheeze signal is highly concentrated in one frequency, f_{50} and f_{90} values of wheeze signals should be close to each other which is depicted for one sample from each class in Figure 3.3.

Thus, f_{50}/f_{90} ratio is expected to be larger for wheeze signals than non-wheeze signals.



(a)



(b)

Figure 3.3. Estimated power spectral densities of (a) Wheeze sample (b) Non-wheeze sample

The histograms of f_{50}/f_{90} ratio for both classes are given in Figure 3.4. The Figure

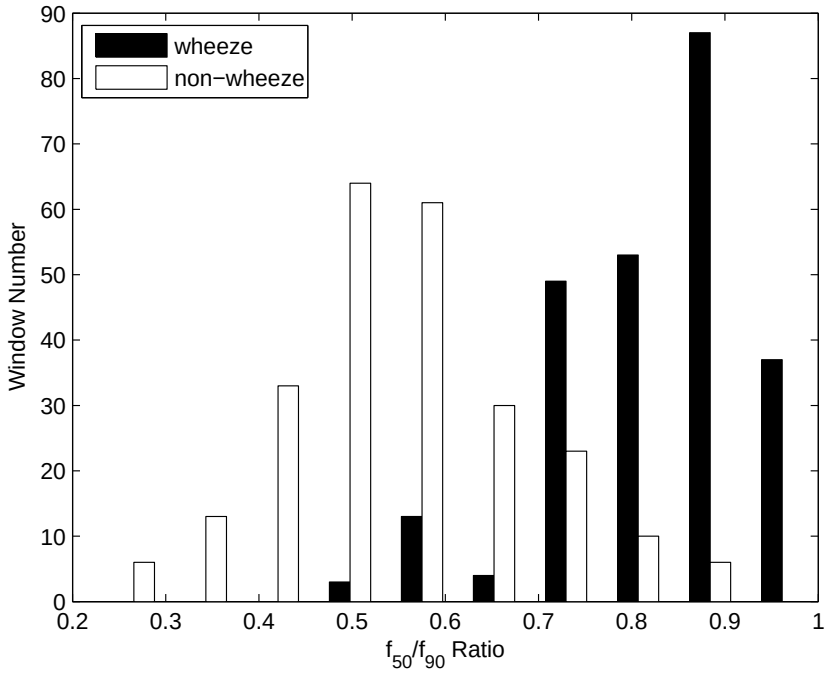


Figure 3.4. Histograms for f_{50}/f_{90} ratio

confirms the expectation that wheeze signals' f_{50}/f_{90} ratios are larger than non-wheeze signals'. Here, Welch spectral estimator is used for psd estimation (see Appendix- A). The length of the window and the point number of FFT are 256, overlapping ratio is 50 % and the window type is Hamming window.

3.3. Mean Crossing Irregularity

It is expected that there should be difference between wheeze and non-wheeze signals which have regular and irregular oscillations, respectively. When the interval between successive mean crossing indexes of the signals are defined as random variables, the deviation from the mean value should be larger for wheeze signals than that of non-wheeze signals. Mathematically, if X denotes the random variable for interval between two successive mean crossing values, the related feature is defined as $\frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}(X)}$, i.e. the ratio of mean value of this variable to its' standard deviation. The histograms of the mean crossing values for both classes are given in the Figure 3.5. As it can be seen from the Figure, the irregularity of mean crossing values of non-wheeze signals

are larger and there is a significant difference between two classes in terms of mean crossing irregularity.

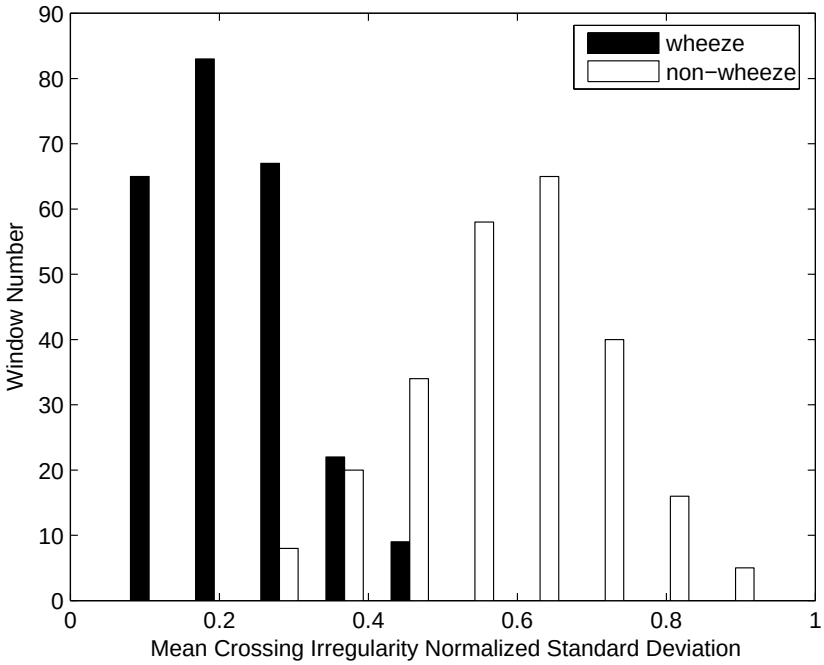


Figure 3.5. Histograms for mean-crossing irregularities

4. MULTI-DIMENSIONAL APPROACH

In this approach, both three-dimensional non-wheeze ($\mathbf{X}_0$) and wheeze ($\mathbf{X}_1$) data are considered as random vectors where 1st, 2nd and 3rd dimensions are the values of Renyi Entropy, f_{50}/f_{90} ratio and mean crossing irregularity, respectively. Moreover, those random vectors are assumed to be Gaussian-distributed with mean vectors $\bar{\mu}_j$ and covariance matrices $\mathbf{\Sigma}_j$, for $j = 0$ (non-wheeze) and $j = 1$ (wheeze). Here, following the standard convention, we let $\mathbf{x}_0 \in \mathcal{X}_0$ and $\mathbf{x}_1 \in \mathcal{X}_1$ denote specific realizations of the corresponding random vectors $\mathbf{X}_0$ and $\mathbf{X}_1$, respectively. If the origins of the data (wheeze or non-wheeze) are unknown, then we let $\mathbf{X}$ denote the three-dimensional random vector that represents the outcome of an observations; in that case, we let $\mathbf{x} \in (\mathcal{X}_0 \cup \mathcal{X}_1)$ denote its specific realization, which may correspond to a "wheeze value" or "non-wheeze value". Moreover, in our experimental setup, we let $\mathbf{x}_{0,i}$ and $\mathbf{x}_{1,i}$ denote the i^{th} realizations from the corresponding classes of wheeze and non-wheeze where $i = \{1, 2, ..., N\}$. These three dimensional data for each class are shown in Figure 4.1. Our next goal is to experimentally verify the Gaussianity assumption of features using several statistical tests which is carried out in Section 4.1.

4.1. Experimental Verification of Gaussianity of Multi-Dimensional Data

Gaussianity assumption may be tested in several ways. In this study, we utilize kurtosis and KL distance measures.

4.1.1. Kurtosis Test

First, kurtosis is used in order to test Gaussianity of each feature, where we utilize empirical estimates of marginal pmf's. The definition for Kurtosis is given next.

Definition 4.1.1 *Kurtosis gives the degree of peakedness of a probability distribution and defined as* $k = \frac{E(X-\mu)^4}{\sigma^4}$ *for a random variable* X.

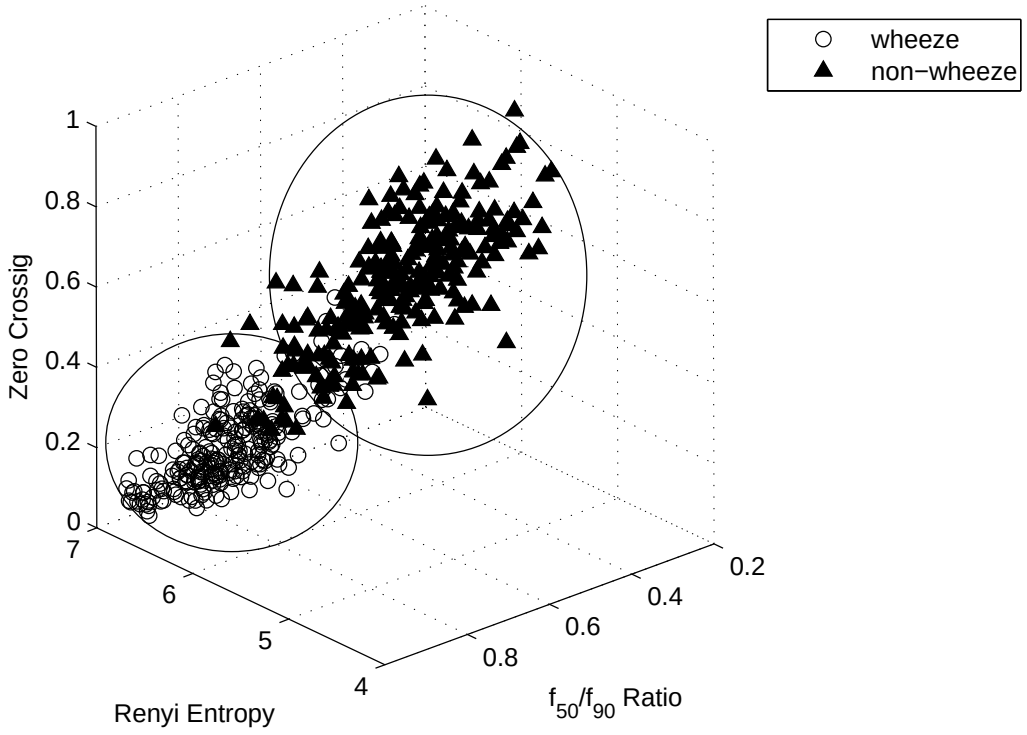


Figure 4.1. Three-dimensional data

Remark 4.1.1 *For a true Gaussian distribution, it is well known that the kurtosis value is exactly equal to* 3*. In practice, if the empirical estimate of the kurtosis is "sufficiently close" to* 3*, this would be a good indicator of the validity of the Gaussian approximation.*

Thus, kurtosis can be used in order to measure the similarity of a given distribution to the Gaussian distribution by considering how close the kurtosis value is to 3. The empirical estimates of kurtosis values of each feature for both classes are given in Table 4.1. According to the results specified in Table 4.1, it can be concluded that p.m.f's of all features both for wheeze and non-wheeze are close to the normal distribution for all practical purposes.

Theorem 4.1.1 *If* $\mathbf{X}$ *is a normal random vector with mean* μ *and covariance matrix* $\mathbf{\Sigma}$*, and* $\mathbf{Z}$ *is an arbitrary random vector with the same dimension as* $\mathbf{X}$*, then, conditioned on* $\mathbf{Z} = \mathbf{z}$ *the result of the dot product* $< \mathbf{Z}, \mathbf{X} > |_{\mathbf{Z}=\mathbf{z}}$ *is also a normal random*

variable with mean $\mathbf{z}^T\mu$ *and covariance matrix* $\mathbf{z}^T\mathbf{\Sigma}\mathbf{z}$.

Table 4.1. Empirical estimates of kurtosis values of each feature

Features	Wheeze	Non-wheeze
Renyi Entropy	2.92	3.46
f_{50}/f_{90} Ratio	3.62	2.90
Mean Crossing Irregularity	2.65	2.70

The experimental results provided in Table 4.1 show that, in practice, it is reasonable to assume that each feature is *marginally* Gaussian-distributed. Note that, this does not necessarily mean that they are jointly-Gaussian-distributed[1] . As such, we also need to experimentally verify the validity of the *joint* Gaussian model of the utilized three features. In order to achieve this task, we use the fact that Gaussianity is preserved under linear transformations (c.f 4.1.1 and [20]); this implies that kurtosis estimate of the distribution of the linear-tranformation-outputs of a Gaussian vector should be sufficiently close to 3 (if joint Gaussianity does not hold, this would not be true). Motivated by this observation, we compute the dot products of three-dimensional vectors, $\{\mathbf{x}_{0,1}, \mathbf{x}_{0,2}, ..., \mathbf{x}_{0,N}\}$ and $\{\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, ..., \mathbf{x}_{1,N}\}$ with pseudo-randomly generated vectors $\{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_M\}$; we subsequently estimate the kurtosis of the distribution of the resulting scalars. Mathematically, if we define the resulting scalars of the dot product between the vectors $\mathbf{z}_m$ and $\mathbf{x}_{j,i}$ as $\gamma_{i,m}^j \stackrel{\triangle}{=} < \mathbf{z}_m, \mathbf{x}_{j,i} >$ for $j = \{0, 1\}$, $i = \{1, ..., N\}$ and $m = \{1, ..., M\}$, then the estimated kurtosis value of the resulting scalars is found via

$$\bar{K}_{m,j} = KurtosisEst(\gamma_{i,m}^j),$$

where the function $KurtosisEst\,(\cdot)$ represents a standard kurtosis estimator. Then, we take averages of these estimated kurtosis values over M such that:

$$\hat{\bar{K}}_j = \frac{1}{M}\sum_{m=1}^{M} \bar{K}_{m,j}$$

[1]In statistical signal processing and probability theory, it is well known that the latter implies the former; however, the converse is not true.

The sample mean of the estimated kurtosis values $\hat{\bar{K}}_0$ and $\hat{\bar{K}}_1$ are found as 3.19 when $N = 245$ and $M = 1000$ for the classes of wheeze and non-wheeze, respectively. Hence we can conclude that the assumption of joint Gaussianity between features is reasonable.

4.1.2. KL Distance Test

In this section, we employ another method to "test" the validity of the Gaussian approximation of the signal samples $\{\mathbf{x}_{j,i}\}$. In particular, we utilize the concept of relative entropy (also known as KL distance) from information theory to achieve this task.

Definition 4.1.2 *The KL distance between two p.m.f's $p(\mathbf{x})$ and $q(\mathbf{x})$ defined over alphabet is defined as [21]*

$$D(p||q) = \sum_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x}) log \frac{p(\mathbf{x})}{q(\mathbf{x})}$$

Remark 4.1.2 *From an information theoretic perspective, the usage of the KL distance to practically verify the validity of an assumed distribution makes sense from a lossless source coding point of view. In order to see that, consider the following setup. Suppose we are given a discrete random variable, $X \in \mathcal{X}$, with the assigned probability measure $p(x)$, then it is well known the fundamental lossless source coding bound is given by $H(X)$, the entropy of X. That is, the expected length of any lossless source code that is designed to compress X is lower bounded by*

$$H(X) = -\sum_{x\in\mathcal{X}} p(x) \log p(x) = E_p\left[\log \frac{1}{p(X)}\right].$$

Note that, the notation $H(p)$ is also used instead of $H(X)$.

Furthermore this bound is asymptotically tight. Next, suppose that the lossless source coder "thinks" that $X \in \mathcal{X}$ is distributed with p.m.f $q(x)$, where $q(\cdot)$ is poten-

tially different from $p(\cdot)$. In that case, it can be shown that, the fundamental lossless source coding bound is replaced by $H(X) + D(p||q)$. That is, the expected length of any source code, that is designed by a source encoder with respect to distribution $q(x)$ is lower bounded by

$$\begin{aligned} E_p\left[\log \frac{1}{q(X)}\right] &= -\sum_{x\in\mathcal{X}} p(x)\log q(x) \\ &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) + \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \\ &= H(X) + D(p||q). \end{aligned}$$

Furthermore, this bound is also asymptotically tight. The reason is intuitively clear: A lossless source coder, that designs a source code with respect to the source distribution $q(x)$ pays the penalty of $D(p||q)$ bits per symbol in the sense of expected codeword length. Normalizing this quantity by the entropy of X, we see that $\frac{D(p||q)}{H(p)}$ is the "relative distribution mismatch penalty" per sample from a lossless source coding point of view [21]. As such, we believe that it can be used to validate a proposed distribution over a sample of realizations.

Next, letting the true pmf be denoted by $p^j(.)$ for $j = \{0, 1\}$, we have $p^j(\mathbf{v}_k) = Pr[\mathbf{X}_j = \mathbf{v}_k]$ where $\mathbf{v}_k = [a_k \ b_k \ c_k]^T \in \mathbb{R}^3$ for $k = \{1, 2, \cdots, K\}$. Based on the observed samples of $\{\mathbf{x}_{j,1}, \mathbf{x}_{j,2}, ..., \mathbf{x}_{j,N}\}$, the empirical estimate $\hat{p}^j(.)$ of $p^j(.)$ is given by

$$\hat{p}^j(\mathbf{v}_k) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\mathbf{x}_{j,i}\in S(\mathbf{v}_k)\}}$$

where $S(\mathbf{v}_k)$ is the sufficiently small cube with the center of $(a_k\ ,\ b_k\ ,\ c_k)$ and in size of $(da \times db \times dc)$ in three-dimensional space of wheeze and non-wheeze.

Next, suppose we follow a parametric approach in the statistical modeling of the signals $\{\mathbf{x}_j\}$ and assume that they are realizations of a three-dimensional Gaussian process, with mean $\bar{\mu}_j$, covariance matrix $\mathbf{\Sigma}_j$. Let the corresponding pdf be denoted

by $q^j(\cdot)$ for $j = 0$ (non-wheeze) and $j = 1$ (wheeze):

$$q^j(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}|\mathbf{\Sigma}_j|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\bar{\mu}_j)^T \mathbf{\Sigma}_j^{-1}(\mathbf{x}-\bar{\mu}_j)\right\}. \tag{4.1}$$

Next, in order to test the validity of this assumption, we "compare" the estimated-discretized version of $q^j(\cdot)$ with $\hat{p}^j(\cdot)$ for $j = 0, 1$. Let $r^j(\cdot)$ denotes the discretized version of $q^j(\cdot)$ for $j = \{0, 1\}$, with the discretization bins $\{S(\mathbf{v}_k)\}$. This implies that

$$\begin{aligned} r^j(\mathbf{v}_k) &= \int_{\mathbf{x}_j \in S(\mathbf{v}_k)} \frac{1}{(2\pi)^{3/2}|\mathbf{\Sigma}_j|^{1/2}} \exp[-\frac{1}{2}(\mathbf{x}-\bar{\mu}_j)^T \mathbf{\Sigma}_j^{-1}(\mathbf{x}-\bar{\mu}_j)] \\ &\approx \frac{1}{(2\pi)^{3/2}|\mathbf{\Sigma}_j|^{1/2}} \exp[-\frac{1}{2}(\mathbf{v}_k-\bar{\mu}_j)^T \mathbf{\Sigma}_j^{-1}(\mathbf{v}_k-\bar{\mu}_j)] dadbdc. \end{aligned}$$

Here, mean values $\bar{\mu}_0$ and $\bar{\mu}_1$ of three-dimensional data are estimated using empirical estimation (i.e., sample means):

$$\hat{\bar{\mu}}_j = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_{j,i} \quad \text{where} \quad j = \{0, 1\}. \tag{4.2}$$

Covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ are estimated via the following formula:

$$\hat{\mathbf{\Sigma}}_j = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{x}_{j,i} - \hat{\bar{\mu}}_j)(\mathbf{x}_{j,i} - \hat{\bar{\mu}}_j)^T \quad \text{where} \quad j = \{0, 1\}. \tag{4.3}$$

Remark 4.1.3 *Equations 4.2 and 4.3 are unbiased ML estimates, $\hat{\bar{\mu}}_j$ and $\hat{\mathbf{\Sigma}}_j$, of $\bar{\mu}_j$ and $\mathbf{\Sigma}_j$, respectively [24].*

Having estimated $(\bar{\mu}_j, \mathbf{\Sigma}_j)$ with $(\hat{\bar{\mu}}_j, \hat{\mathbf{\Sigma}}_j)$ (c.f. 4.2 and 4.3), the empirical estimate $\hat{r}^j(.)$ of $r^j(.)$ is then given by

$$\hat{r}^j(\mathbf{v}_k) \approx \frac{1}{(2\pi)^{3/2}|\hat{\mathbf{\Sigma}}_j|^{1/2}} \exp[-\frac{1}{2}(\mathbf{v}_k-\hat{\bar{\mu}}_j)^T \hat{\mathbf{\Sigma}}_j^{-1}(\mathbf{v}_k-\hat{\bar{\mu}}_j)] dadbdc.$$

Then, the estimated "normalized" KL distance between $\hat{p}^j$ and $\hat{r}^j$, i.e. $\frac{D(\hat{p}^j||\hat{r}^j)}{H(\hat{p}^j)}$, are found as 0.07 and 0.09 over 125 cubes, i.e. $K = 125$ for $j = 1$ (wheeze) and $j = 0$

(non-wheeze), respectively.

Hence, according to estimated kurtosis and "normalized" KL distance values, the distributions of three-dimensional data can be assumed as Gaussian. In practice, we assumed that the three-dimensional data values are Gaussian distributed with means $\hat{\bar{\mu}}_0$, $\hat{\bar{\mu}}_1$ and covariance matrices $\hat{\mathbf{\Sigma}}_0$, $\hat{\mathbf{\Sigma}}_1$ for non-wheeze and wheeze, respectively.

4.2. Detection Theoretic Analysis

For a given pdf's in equation 4.1 of three-dimensional data, the log-likelihood ratio between hypothesis H_0 (hypothesis of being non-wheeze) and H_1 (hypothesis of being wheeze) is given by

$$\begin{aligned}
\log \mathrm{L}(\mathbf{x}) &= \log\frac{q_1(\mathbf{x})}{q_0(\mathbf{x})} \\
&= \log\left(\frac{|\mathbf{\Sigma}_0|^{1/2}}{|\mathbf{\Sigma}_1|^{1/2}}\exp\left\{\frac{1}{2}\left[(\mathbf{x}-\bar{\mu}_0)^T\mathbf{\Sigma}_0^{-1}(\mathbf{x}-\bar{\mu}_0)-(\mathbf{x}-\bar{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{x}-\bar{\mu}_1)\right]\right\}\right) \\
&= \frac{1}{2}\log\frac{|\mathbf{\Sigma}_0|}{|\mathbf{\Sigma}_1|}+\frac{1}{2}\left[(\mathbf{x}-\bar{\mu}_0)^T\mathbf{\Sigma}_0^{-1}(\mathbf{x}-\bar{\mu}_0)-(\mathbf{x}-\bar{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\mathbf{x}-\bar{\mu}_1)\right].
\end{aligned}$$

Then, the decision rule between two hypothesis is defined by

$$\begin{aligned}
H_0: &\quad \mathbf{x}\sim\mathcal{N}(\bar{\mu}_0,\mathbf{\Sigma}_0) \quad \text{if} \quad \log \mathrm{L}(\mathbf{x})<\tau \\
H_1: &\quad \mathbf{x}\sim\mathcal{N}(\bar{\mu}_1,\mathbf{\Sigma}_1) \quad \text{if} \quad \log \mathrm{L}(\mathbf{x})>\tau.
\end{aligned} \tag{4.4}$$

To evaluate the statistical performance of the algorithm, the probability of false alarm $\mathrm{P}_F(\cdot)$ (probability of saying wheeze for a given non-wheeze window) and the probability of miss $\mathrm{P}_M(\cdot)$ (probability of saying non-wheeze for a given wheeze window) are defined as

$$\begin{aligned}
\mathrm{P}_F(\tau) &= \int_{\mathcal{X}}\mathrm{P}(\log \mathrm{L}(\mathbf{x})>\tau|H_0)d\mathbf{x}=\int_{\mathcal{X}}\mathrm{P}(\log \mathrm{L}(\mathbf{x})>\tau|\mathbf{x}\in\mathcal{X}_0)d\mathbf{x} \\
\mathrm{P}_M(\tau) &= \int_{\mathcal{X}}\mathrm{P}(\log \mathrm{L}(\mathbf{x})<\tau|H_1)d\mathbf{x}=\int_{\mathcal{X}}\mathrm{P}(\log \mathrm{L}(\mathbf{x})<\tau|\mathbf{x}\in\mathcal{X}_1)d\mathbf{x}.
\end{aligned}$$

Now, finding the optimum value of τ is another problem to be considered. The optimum value of the threshold τ should be selected in the sense of minimizing error, i.e. $\mathrm{P}_F(\cdot)$ and $\mathrm{P}_M(\cdot)$. However, since it is a controversial concept that which one of the error probability has higher importance and the prior probabilities are unknown, ROC curve for sufficiently many values of threshold τ are utilized. Indeed, for any value of τ the total probability of error can be calculated by

$$\mathrm{P}_e(\tau) = \mathrm{P}(H_0)\mathrm{P}_F(\tau) + \mathrm{P}(H_1)\mathrm{P}_M(\tau) \tag{4.5}$$

where $\mathrm{P}(H_0)$ and $\mathrm{P}(H_1)$ are prior probabilities of the hypothesis H_0 and H_1, respectively. The minimum value of the probability of error is denoted as $\mathrm{P}_e(\tau^*)$ for optimum value of τ^*.

4.3. Results for Multi-dimensional Approach

4.3.1. Theoretical Results

Using equations 4.2 and 4.3 as the estimations of mean vectors $\bar{\mu}_j$ and covariance matrices $\mathbf{\Sigma}_j$ of three-dimensional data, the estimated log-likelihood ratio can be written as:

$$\log \hat{L}(\mathbf{x}) = \frac{1}{2}\log\frac{|\hat{\mathbf{\Sigma}}_0|}{|\hat{\mathbf{\Sigma}}_1|} + \frac{1}{2}\left[(\mathbf{x}-\hat{\bar{\mu}}_0)^T\hat{\mathbf{\Sigma}}_0^{-1}(\mathbf{x}-\hat{\bar{\mu}}_0) - (\mathbf{x}-\hat{\bar{\mu}}_1)^T\hat{\mathbf{\Sigma}}_1^{-1}(\mathbf{x}-\hat{\bar{\mu}}_1)\right]. \tag{4.6}$$

Then, the decision rule in equation 4.4 turns out to be:

$$\begin{aligned} H_0 :&\quad \mathbf{x} \sim \mathcal{N}(\hat{\bar{\mu}}_0, \hat{\mathbf{\Sigma}}_0) \quad \text{if} \quad \log \hat{L}(\mathbf{x}) < \tau \\ H_1 :&\quad \mathbf{x} \sim \mathcal{N}(\hat{\bar{\mu}}_1, \hat{\mathbf{\Sigma}}_1) \quad \text{if} \quad \log \hat{L}(\mathbf{x}) > \tau \end{aligned} \tag{4.7}$$

with estimated error probabilities

$$\begin{aligned}\hat{\mathrm{P}}_F(\tau) &= \int_{\mathcal{X}} \mathrm{P}(\log \hat{L}(\mathbf{x}) > \tau | H_0) d\mathbf{x} = \int_{\mathcal{X}} \mathrm{P}(\log \hat{L}(\mathbf{x}) > \tau | \mathbf{x} \in \mathcal{X}_0) d\mathbf{x} \\ \hat{\mathrm{P}}_M(\tau) &= \int_{\mathcal{X}} \mathrm{P}(\log \hat{L}(\mathbf{x}) < \tau | H_1) d\mathbf{x} = \int_{\mathcal{X}} \mathrm{P}(\log \hat{L}(\mathbf{x}) < \tau | \mathbf{x} \in \mathcal{X}_1) d\mathbf{x}\end{aligned}$$

which gives the estimated total probability of error for any value of τ as:

$$\hat{\mathrm{P}}_e(\tau) = \mathrm{P}(H_0)\hat{\mathrm{P}}_F(\tau) + \mathrm{P}(H_1)\hat{\mathrm{P}}_M(\tau) \tag{4.8}$$

Then, ROC curve is drawn in Figure 4.2 for estimated error probabilities. From Figure

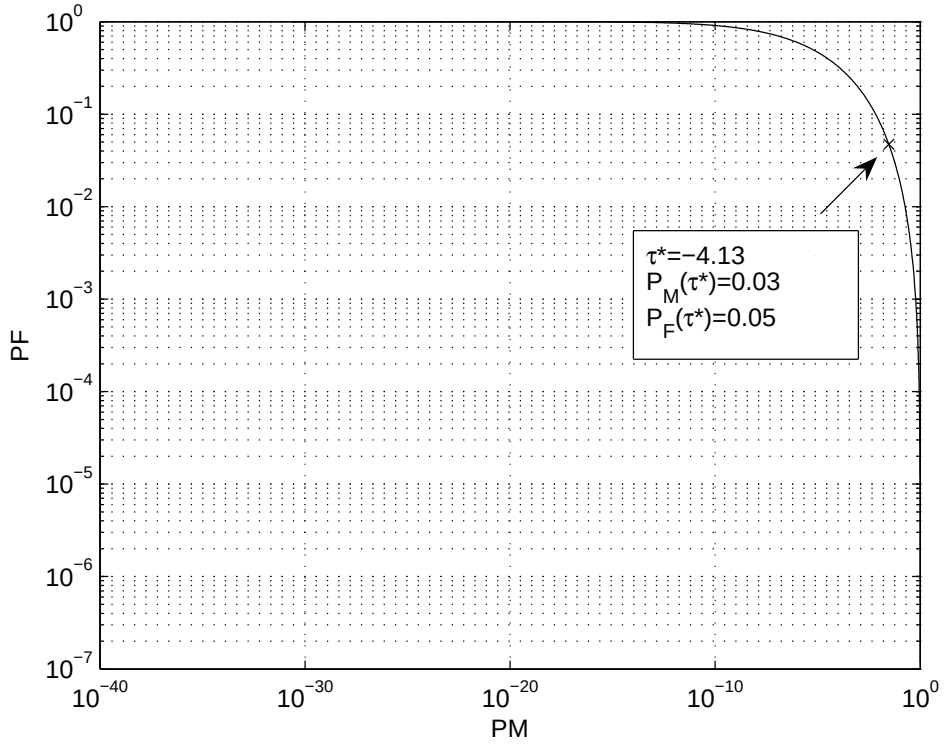


Figure 4.2. Theoretical ROC curve for multi-dimensional approach

4.2, the minimum probability of error $\hat{\mathrm{P}}_e(\tau^*)$ is found as 0.04 where $\hat{\mathrm{P}}_M(\tau^*) = 0.03$ and $\hat{\mathrm{P}}_F(\tau^*) = 0.05$ for $\tau^* = -4.13$ under assumption of equal priors ($\mathrm{P}(H_1) = \mathrm{P}(H_0) = 0.5$).

4.3.2. Empirical Results

The decision rule in equation 4.7 is applied for the existent three-dimensional data of wheeze and non-wheeze ($N = 245$ for each). Then, the number of false positives (number of non-wheeze windows which are said to be wheeze) and false negatives (number of wheeze windows which are said to be non-wheeze) for some values of τ are given in Table 4.2. From the table, it can be seen that the minimum number of the sum

Table 4.2. Number of False Positives and False Negatives for different τ values for multi-dimensional approach

τ	Number of False Positives	Number of False Negatives
-6.7	0	157
-6.4	0	141
-4.7	1	79
-3.3	3	43
-1.8	8	21
-0.76	16	11
-0.48	18	9
0.08	19	5
0.36	21	5
0.94	28	3
4.11	59	0
6.16	70	0

of false negatives and false positives are found as 24 where the number of false positives is 19 and the number of false negatives is 5, respectively. Regarding the total number of false positives and false negatives as the performance measure of the approach, the percentage error rate over the data set is found to be $24/490 = \%4.9$.

In order to see the performance of the method in more detail, the operational ROC curve is used. For this purpose, empirical estimated error probabilities are defined

as:

$$\hat{\mathrm{P}}_{M,emp}(\tau) = \frac{\sum_{i=1}^{N} \mathbf{1}_{\log \hat{L}(\mathbf{x}_{1,i})<\tau}}{N} = \frac{\text{Number of false negatives}}{\text{Total number of wheeze windows}}$$
$$\hat{\mathrm{P}}_{F,emp}(\tau) = \frac{\sum_{i=1}^{N} \mathbf{1}_{\log \hat{L}(\mathbf{x}_{0,i})>\tau}}{N} = \frac{\text{Number of false positives}}{\text{Total number of non-wheeze windows}}$$

and the estimated total error probability is calculated via:

$$\hat{\mathrm{P}}_{e,emp}(\tau) = \mathrm{P}(H_0)\hat{\mathrm{P}}_{F,emp} + \mathrm{P}(H_1)\hat{\mathrm{P}}_{M,emp} \tag{4.9}$$

Then, operational ROC curve is drawn in Figure 4.3 using the empirical estimated error probabilities for different τ values. According to the Figure, optimum value of threshold

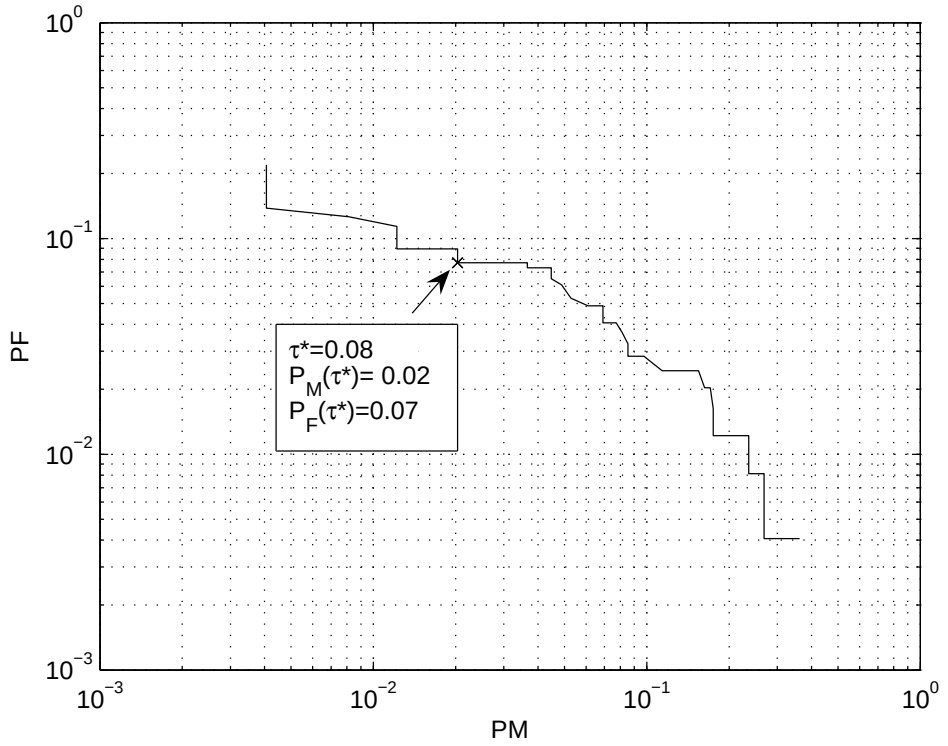


Figure 4.3. Empirical ROC curve for multi-dimensional approach

τ^* is found to be 0.08 where $\hat{\mathrm{P}}_{M,emp}(\tau) + \hat{\mathrm{P}}_{F,emp}(\tau)$ is minimized and the empirical estimated error probabilities for this value of threshold are found as $\mathrm{P}_{M,emp}(\tau^*) = 0.02$ and $\mathrm{P}_{F,emp}(\tau^*) = 0.07$ which results with total probability of error 0.045 under equal priors assumption.

5. ONE-DIMENSIONAL APPROACH

5.1. Dimensionality Reduction Using FDA

Fisher Discrimination method is applied in order to represent three-dimensional data in one-dimensional space where the two classes can be separated best. Recall that we have a set of three-dimensional data $\{ \mathbf{x}_{j,i} \in \mathcal{X}_j\}$ from wheeze $(j = 1)$ and non-wheeze $(j = 0)$ classes where $i \in \{ 1, 2, ..., N\}$. When the dot product of the data set with three-dimensional vector $\mathbf{w} \in \mathbb{R}^3$ is taken $\left(y_{j,i} =< \mathbf{w}, \mathbf{x}_{j,i} >= \mathbf{w}^T\mathbf{x}_{j,i}\right)$, the result will be the scalars $\{ y_{j,i} \in \mathcal{Y}_j\}$ which are the specific realizations of the random variables Y_j for $j = \{0, 1\}$. The scalars $\{ y_{j,i}\}$ are the projected values of the three-dimensional data $\{ \mathbf{x}_{j,i}\}$ onto the line in the direction of the vector $\mathbf{w}$. Hence the magnitude of the vector $\mathbf{w}$ is of no important but the direction is.

Now, the question is how to find the best $\mathbf{w}$ vector for the accurate classification. The difference between the sample averages of the projected values of two classes is one of the measure desired to be increased for good separation. However, this measure should be defined with standard deviations for each classes. Hence, the distance measure between sample averages relative to some measure of standard deviation can be used as the cost function needed to be maximized over $\mathbf{w} \in \mathbb{R}^3$ vectors.

Recalling the sample mean of three-dimensional data (c.f. 4.2), the sample means of the projected values can be written as

$$\begin{aligned} \hat{\mu}_j &= \frac{1}{N}\sum_{i=1}^{N} y_{j,i} \\ &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{w}^T\mathbf{x}_{j,i} \\ &= \mathbf{w}^T\hat{\bar{\mu}}_j \end{aligned}$$

Then, the distance between the sample averages of the projected values is

$$|\hat{\mu}_1 - \hat{\mu}_0| = |\mathbf{w}^T(\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0)|.$$

Considering the standard deviation of the classes, the term *scatter* of the projected values of each class is defined

$$\hat{s}_j^2 = \sum_{i=1}^{N}(y_{j,i} - \hat{\mu}_j)^2$$

which is merely the scaled version of the sample variances. Then, the total *within-class scatter* of the projected values is defined as $\hat{s}_0^2 + \hat{s}_1^2$. Hence the cost function for *Fisher Linear Discriminant* can be written as

$$J(\mathbf{w}) = \frac{|\hat{\mu}_1 - \hat{\mu}_0|^2}{\hat{s}_0^2 + \hat{s}_1^2}$$

which is desired to be maximum. In order to get $J(\cdot)$ as an explicit function of $\mathbf{w}$, the scatter matrices $\mathbf{S}_j$ and $\mathbf{S}_w$ are defined

$$\begin{aligned} \mathbf{S}_j &= \sum_{i=1}^{N}(\mathbf{x}_{j,i} - \hat{\bar{\mu}}_j)^T \\ \mathbf{S}_W &= \mathbf{S}_0 + \mathbf{S}_1 \end{aligned}$$

Note that since the *within-class scatter* matrix $\mathbf{S}_W$ is scaled version of sample covariance matrix of the whole data, it is symmetric, positive definite and nonsingular. Next, it can be written that

$$\begin{aligned} \hat{s}_j^2 &= \sum_{i=1}^{N}(\mathbf{w}^T\mathbf{x}_{j,i} - \mathbf{w}^T\hat{\bar{\mu}}_j)^2 \\ &= \sum_{i=1}^{N}\mathbf{w}^T(\mathbf{x}_{j,i} - \hat{\bar{\mu}}_j)(\mathbf{x}_{j,i} - \hat{\bar{\mu}}_j)^T\mathbf{w} \\ &= \mathbf{w}^T\mathbf{S}_j\mathbf{w}. \end{aligned}$$

Hence, the *within-class scatter* of the projected values can be written as

$$\hat{s}_0 + \hat{s}_1 = \mathbf{w}^T \mathbf{S}_W \mathbf{w}.$$

The distance between sample averages of the projected values also can be written as

$$\begin{aligned}(\hat{\mu}_1 - \hat{\mu}_0)^2 &= (\mathbf{w}^T\hat{\bar{\mu}}_1 - \mathbf{w}^T\hat{\bar{\mu}}_0)^2 \\ &= \mathbf{w}^T(\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0)(\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0)^T\mathbf{w} \\ &= \mathbf{w}^T\mathbf{S}_B\mathbf{w}\end{aligned}$$

where

$$\mathbf{S}_B = (\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0)(\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0)^T$$

Finally, the criterion function can be written in terms of $\mathbf{w}$, $\mathbf{S}_B$ and $\mathbf{S}_W$ as

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}. \tag{5.1}$$

In order to maximize Eqn. 5.1, the gradient of the cost function over $\mathbf{w}$ is required to be equal to zero for the optimum vector $\mathbf{w}^* \in \mathbb{R}^3$ as in the following equation

$$\begin{aligned}\nabla_{\mathbf{w}} J(\mathbf{w}^*) &= \frac{2\mathbf{S}_B\mathbf{w}^*(\mathbf{w}^{*T}\mathbf{S}_W\mathbf{w}^*) - 2\mathbf{S}_W\mathbf{w}^*(\mathbf{w}^{*T}\mathbf{S}_B\mathbf{w}^*)}{(\mathbf{w}^{*T}\mathbf{S}_W\mathbf{w}^*)^2} \\ &= 0 \\ &\Rightarrow \mathbf{S}_B\mathbf{w}^*(\mathbf{w}^{*T}\mathbf{S}_W\mathbf{w}^*) = \mathbf{S}_W\mathbf{w}^*(\mathbf{w}^{*T}\mathbf{S}_B\mathbf{w}^*) \\ &\Rightarrow \mathbf{S}_B\mathbf{w}^* = \mathbf{S}_W\mathbf{w}^* \underbrace{\frac{(\mathbf{w}^{*T}\mathbf{S}_B\mathbf{w}^*)}{(\mathbf{w}^{*T}\mathbf{S}_W\mathbf{w}^*)}}_{\lambda} \\ &\Rightarrow \mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w}^* = \lambda\mathbf{w}^*\end{aligned}$$

Since $\mathbf{S}_B\mathbf{w}$ is in the direction of $\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0$ and the magnitude of the vector $\mathbf{w}$ is not

significant, the optimum value $\mathbf{w}^*$ for $\mathbf{w}$ can be written as

$$\mathbf{w}^* = \mathbf{S}_W^{-1}(\hat{\bar{\mu}}_1 - \hat{\bar{\mu}}_0)$$

[22]. For this $\mathbf{w}^*$ vector, the projection values of both classes are given in Figure 5.1. Before applying the decision rule we need to characterize the density of the one-

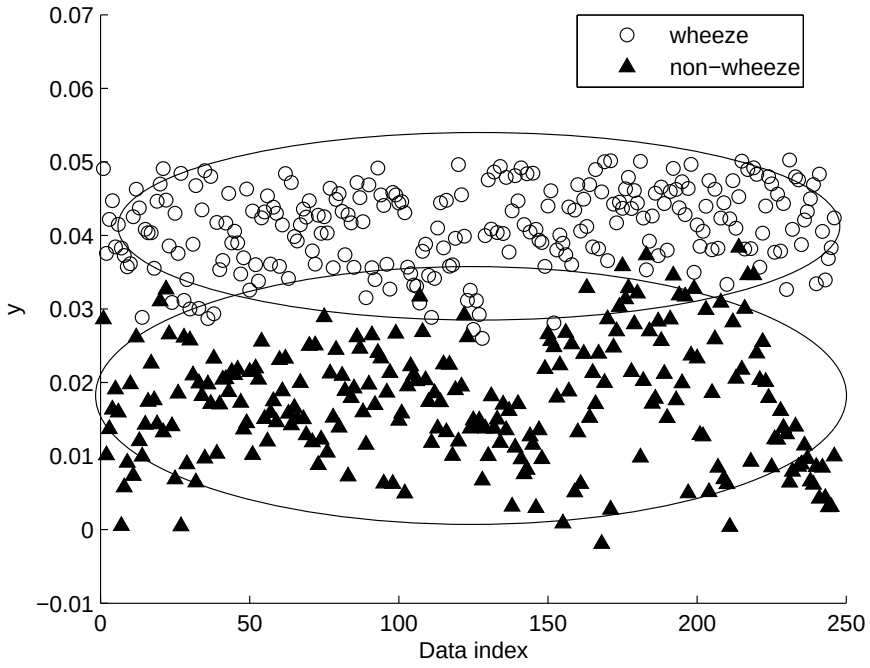


Figure 5.1. One-dimensional data after FDA

dimensional data in Figure 5.1. The term "one-dimensional data" will be used interchangeably with the term "projection values" through the thesis. In this approach, we assumed that these one-dimensional data of non-wheeze and wheeze vectors are distributed with means μ_0, μ_1 and variances σ_0^2, σ_1^2, respectively. However, again, the Gaussianity test is required to verify the Gaussianity assumption of projected values. In order to measure the similarity of the distribution of projection values with Gaussian distribution as kurtosis and normalized KL distance measures are utilized as in Section 4.1.

5.2. Experimental Verification of Gaussianity of One-Dimensional Data

Gaussianity assumption of one-dimensional data is verified as three-dimensional data in multi-dimensional approach. For this purpose, again, kurtosis and normalized KL distance measures are used.

5.2.1. Kurtosis Test

Recalling the definition of kurtosis in section 4.1.1, estimated kurtosis values $\hat{K}_j = KurtosisEst(y_1^j, ..., y_N^j)$ of the one-dimensional data $\{y_{j,i}\}$ where $j = \{0, 1\}$ and $i = \{1, 2, \cdots, N\}$ are found as $\hat{K}_1 = 2.50$ and $\hat{K}_0 = 2.53$ for wheeze ($j = 1$) and non-wheeze ($j = 0$), respectively. From the results of the estimated kurtosis values, it can be concluded that the Gaussianity assumption of one-dimensional data is reasonable. These results are expected because it is known that the linear transformation of Gaussian random vectors are also normal Gaussian according to Theorem 4.1.1. Therefore, since the Gaussianity of three-dimensional data $\{\mathbf{x}_{j,i}\}$ for $j = \{0, 1\}$ and $i = \{1, 2, \cdots, N\}$ is verified in section 4.1, the one-dimensional data which are just the linear transformations of multi-dimensional data should also be Gaussian which is verified by the estimated kurtosis values.

5.2.2. KL Distance Test

We denote $p^j(\cdot)$ as the true p.m.f of one-dimensional data and $q^j(\cdot)$ as its' pdf approximation under Gaussian assumption. For true p.m.f, we have $p^j(a_k) = Pr(Y_j = a_k)$ for $a_k \in \mathbb{R}$ and $k = \{1, 2, \cdots, K\}$. Next, empirical estimate $\hat{p}^j(\cdot)$ of p.m.f $p^j(\cdot)$ based on the observed one-dimensional data $\{y_{j,i}\}$ is given by

$$\hat{p}^j(a_k) = \sum_{i=1}^{N} \frac{\mathbf{1}_{y_{j,i} \in S(a_k)}}{N}$$

where $S(a_k) = \{a_k - da < a_k < a_k + da\}$ is the sufficiently small region with the center of a_k in one-dimensional space of wheeze and non-wheeze.

Next, under Gaussianity assumption, we have Gaussian pdf's with means μ_j and variances σ_j^2 of one-dimensional data for wheeze ($j = 1$) and non-wheeze ($j = 0$) such that

$$q^j(y) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(y-\mu_j)^2}{2\sigma_j^2}\right\}. \tag{5.2}$$

In order to test validity of this assumption, we utilize normalized KL distance between estimated p.m.f and discretized version of pdf of one-dimensional data as in section 4.1.2. Let $r^j(\cdot)$ denote the discretized pdf for $j = \{0, 1\}$ with bins $\{S(a_k)\}$ which gives

$$\begin{aligned} r^j(a_k) &= \int_{y \in S(a_k)} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(y-\mu_j)^2}{2\sigma_j^2}\right\} dy \\ &= Q\left(\frac{y - da - \mu_j}{\sigma_j}\right) - Q\left(\frac{y + da - \mu_j}{\sigma_j}\right). \end{aligned}$$

See Appendix- B for Q-function. Here, the mean values μ_0, μ_1 and variances σ_0, σ_1 of the one-dimensional data are estimated by

$$\hat{\mu}_j = \frac{1}{N}\sum_{i=1}^{N} y_{j,i} \tag{5.3}$$

$$\hat{\sigma}_j^2 = \frac{1}{N-1}\sum_{i=1}^{N} (y_{j,i} - \hat{\mu}_j)^2 \tag{5.4}$$

where $j = \{0, 1\}$. According to the estimated parameters $\hat{\mu}_j$, $\hat{\sigma}_j^2$ and Gaussian assumption of one-dimensional data, the estimated pdf's are given in Figure 5.2. Moreover, the estimated dicretized pdf is written by:

$$\hat{r}^j(a_k) = Q\left(\frac{y - da - \hat{\mu}_j}{\hat{\sigma}_j}\right) - Q\left(\frac{y + da - \hat{\mu}_j}{\hat{\sigma}_j}\right)$$

Then, normalized KL distances $\left(\frac{D(\hat{p}^j||\hat{r}^j)}{H(\hat{p}^j)}\right)$ between $\hat{p}^j(\cdot)$ and $\hat{r}^j(\cdot)$ both for wheeze ($j = 1$) and non-wheeze ($j = 0$) are found as 0.06 and 0.04 over 128 sub-regions, respectively.

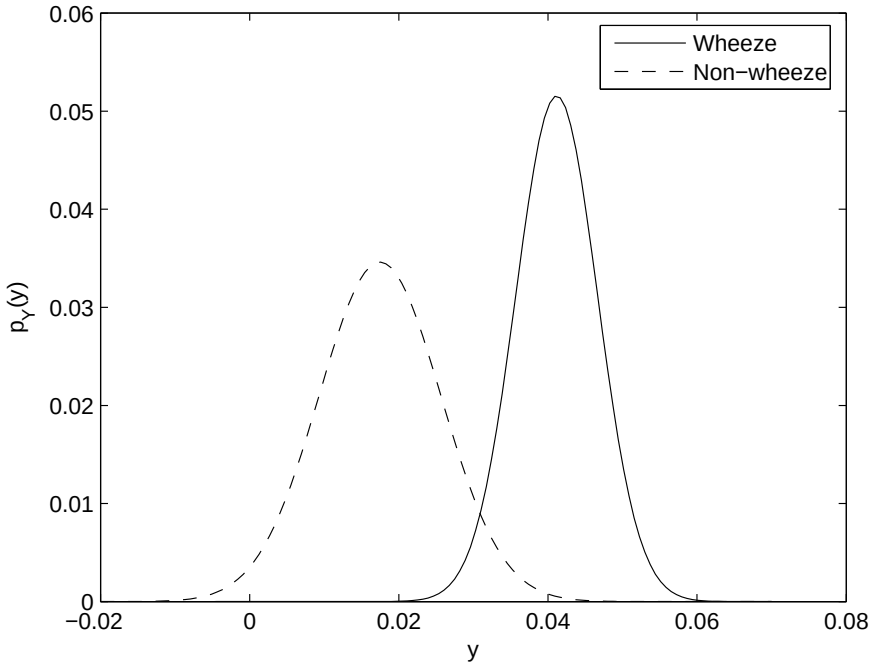


Figure 5.2. Estimated Distributions of one-dimensional data under Gaussianity assumption

Hence, according to estimated kurtosis and normalized KL distance values, the distributions of one-dimensional data can be assumed as Gaussian. In practice, we assumed that the one-dimensional data are distributed with means $\hat{\mu}_0$, $\hat{\mu}_1$ and variances $\hat{\sigma}_0^2$, $\hat{\sigma}_1^2$ for non-wheeze and wheeze, respectively.

5.3. Detection Theoretic Analysis

The decision rule between hypothesis H_0 (hypothesis of being non-wheeze) and H_1 (hypothesis of being wheeze) for one-dimensional data is defined as

$$\begin{aligned} H_0: &\quad y \sim \mathcal{N}(\mu_0, \sigma_0) \quad \text{if} \quad \log \mathrm{L}(y) < \tau \\ H_1: &\quad y \sim \mathcal{N}(\mu_1, \sigma_1) \quad \text{if} \quad \log \mathrm{L}(y) > \tau \end{aligned} \tag{5.5}$$

where the log-likelihood ratio ($\log \mathrm{L}(y)$) for a given pdf's in equation 5.2 is:

$$\begin{aligned}
\log \mathrm{L}(y) &= \log \frac{q_1(y)}{q_0(y)} \\
&= \log \left(\frac{\sigma_0}{\sigma_1} \exp \left\{ \frac{1}{2} \frac{(y-\mu_0)^2}{\sigma_0^2} - \frac{1}{2} \frac{(y-\mu_1)^2}{\sigma_1^2} \right\} \right) \\
&= \log \frac{\sigma_0}{\sigma_1} + \frac{1}{2} \frac{(y-\mu_0)^2}{\sigma_0^2} - \frac{1}{2} \frac{(y-\mu_1)^2}{\sigma_1^2}.
\end{aligned}$$

Here, it should be noted that the decision rule, i.e. log-likelihood ratio can also be written as an explicit function of $\mathbf{x}$ if the results $\sigma_0^2 = \mathbf{w}^T \mathbf{\Sigma}_0 \mathbf{w}$, $\sigma_1^2 = \mathbf{w}^T \mathbf{\Sigma}_1 \mathbf{w}$, $\mu_0 = \mathbf{w}^T \bar{\mu}_0$ and $\mu_1 = \mathbf{w}^T \bar{\mu}_1$ from Theorem 4.1.1 are used. Then, the log-likelihood ratio can be rewritten as

$$\begin{aligned}
\log \mathrm{L}(\mathbf{x}) &= \log \frac{\sqrt{\mathbf{w}^T \Sigma_0 \mathbf{w}}}{\sqrt{\mathbf{w}^T \Sigma_1 \mathbf{w}}} + \frac{1}{2} \frac{(\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \bar{\mu}_0)^2}{\mathbf{w}^T \mathbf{\Sigma}_0 \mathbf{w}} - \frac{1}{2} \frac{(\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \bar{\mu}_1)^2}{\mathbf{w}^T \mathbf{\Sigma}_1 \mathbf{w}} \\
&= \log \frac{\sqrt{\mathbf{w}^T \Sigma_0 \mathbf{w}}}{\sqrt{\mathbf{w}^T \Sigma_1 \mathbf{w}}} + \frac{1}{2} (\mathbf{x} - \bar{\mu}_0)^T \mathbf{w} (\mathbf{w}^T \mathbf{\Sigma}_0 \mathbf{w})^{-1} \mathbf{w}^T (\mathbf{x} - \bar{\mu}_0) \\
&- \frac{1}{2} (\mathbf{x} - \bar{\mu}_1)^2 \mathbf{w} (\mathbf{w}^T \mathbf{\Sigma}_1 \mathbf{w})^{-1} \mathbf{w}^T (\mathbf{x} - \bar{\mu}_1)
\end{aligned}$$

which contains quadratic and linear terms in $\mathbf{x}$. As in multi-dimensional case, probability of false alarm P_F (probability of saying that the observed one-dimensional data y belongs to wheeze although it is from non-wheeze class) and the probability of miss P_M (probability of saying that the observed one-dimensional data y belongs to non-wheeze although it is from wheeze class) are defined as:

$$\begin{aligned}
\mathrm{P}_M(\tau) &= \int_{\mathcal{Y}} \mathrm{P}(\log \mathrm{L}(y) > \tau | H_0) dy = \int_{\mathcal{Y}} \mathrm{P}(\log \mathrm{L}(y) > \tau | y \in \mathcal{Y}_0) dy \\
\mathrm{P}_F(\tau) &= \int_{\mathcal{Y}} \mathrm{P}(\log \mathrm{L}(y) < \tau | H_1) dy = \int_{\mathcal{Y}} \mathrm{P}(\log \mathrm{L}(y) < \tau | y \in \mathcal{Y}_1) dy.
\end{aligned}$$

in order to evaluate the performance of the algorithm. The equation 4.5 can be used in order to compute probability of error.

5.4. Results for One-dimensional Approach

5.4.1. Theoretical Results

Using, equations 5.3 and 5.4 for means and variances of the one-dimensional data, the estimated log-likelihood ratio can be written as

$$\log \hat{L}(y) = \log\frac{\hat{\sigma}_0}{\hat{\sigma}_1} + \frac{1}{2}\frac{(y-\hat{\mu}_0)^2}{\hat{\sigma}_0^2} - \frac{1}{2}\frac{(y-\hat{\mu}_1)^2}{\hat{\sigma}_1^2}.$$

Then, the decision rule in equation 5.5 becomes:

$$\begin{aligned} H_0: &\quad y \sim \mathcal{N}(\hat{\mu}_0, \hat{\sigma}_0) \quad \text{if} \quad \log \hat{L}(y) < \tau \\ H_1: &\quad y \sim \mathcal{N}(\hat{\mu}_1, \hat{\sigma}_1) \quad \text{if} \quad \log \hat{L}(y) > \tau \end{aligned} \tag{5.6}$$

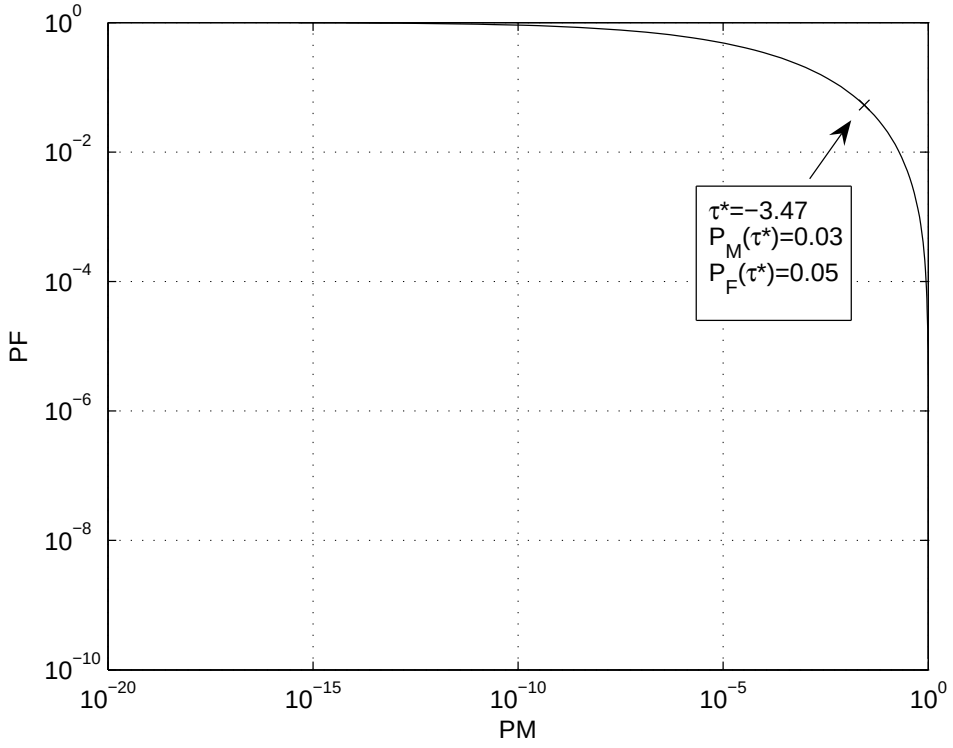


Figure 5.3. Theoretical ROC curve for one-dimensional approach

with the estimated error probabilities

$$\begin{aligned}\hat{\mathrm{P}}_M(\tau) &= \int_{\mathcal{Y}} \mathrm{P}(\log \hat{L}(y) > \tau | H_0) dy = \int_{\mathcal{Y}} \mathrm{P}(\log \hat{L}(y) > \tau | y \in \mathcal{Y}_0) dy \\ \hat{\mathrm{P}}_F(\tau) &= \int_{\mathcal{Y}} \mathrm{P}(\log \hat{L}(y) < \tau | H_1) dy = \int_{\mathcal{Y}} \mathrm{P}(\log \hat{L}(y) < \tau | y \in \mathcal{Y}_1) dy.\end{aligned}$$

The estimated total probability of error $\left(\hat{\mathrm{P}}_e(\tau)\right)$ can be found via the equation 4.8. Then, ROC curve of the decision rule in equation 5.6 for sufficiently many values of τ is shown in Figure 5.3. According to the Figure 5.3, the estimated minimum probability of error $\left(\hat{\mathrm{P}}_e(\tau^*)\right)$ under the equal prior assumption is found as 0.04 where $\hat{\mathrm{P}}_M(\tau^*) = 0.03$ and $\hat{\mathrm{P}}_F(\tau^*) = 0.05$ for the optimum threshold $\tau^* = -3.47$.

5.4.2. Empirical Results

In order to test the experimental performance of the approach, the decision rule in equation 5.6 is applied for the existent one-dimensional data of wheeze and non-wheeze ($N = 245$ for each). For different values of threshold τ, number of false negatives (number of wheeze windows which are said to be non-wheeze) and number of false positives (number of non-wheeze windows which are said to be wheeze) are given in Table 5.1. From the table, it can be seen that, the minimum of the sum of errors is found as 25 where the number of false positive is 22 and false negative is 3 when the value of τ is 0.8. As in Section 4.3.2, the minimum value of the sum of false negatives and false positives is used as a measure of the performance of the method. Hence, the percentage error rate of the method is calculated as $25/490 = \%5.3$. However, the table is not able to reflect the performance of the method in general. Then, ROC curve is used to see the behavior of the method for sufficiently many different threshold values τ. For this purpose, the empirical estimated error probabilities are defined as:

$$\begin{aligned}\hat{\mathrm{P}}_{M,emp}(\tau) &= \frac{\sum_{i=1}^{N} \mathbf{1}_{\log \hat{L}(y_{1,i}) < \tau}}{N} = \frac{\text{Number of false negatives}}{\text{Total number of wheeze windows}} \\ \hat{\mathrm{P}}_{F,emp}(\tau) &= \frac{\sum_{i=1}^{N} \mathbf{1}_{\log \hat{L}(y_{0,i}) > \tau}}{N} = \frac{\text{Number of false positives}}{\text{Total number of non-wheeze windows}}.\end{aligned}$$

Table 5.1. Number of False Positives and False Negatives for different τ values for one-dimensional approach

τ	Number of False Positive	Number of False Negative
-2.3	3	31
-1.2	10	16
-0.7	14	15
-0.3	11	17
0.4	20	8
0.8	22	3
1.2	27	2
2.3	36	1
2.9	46	0
3.9	57	0

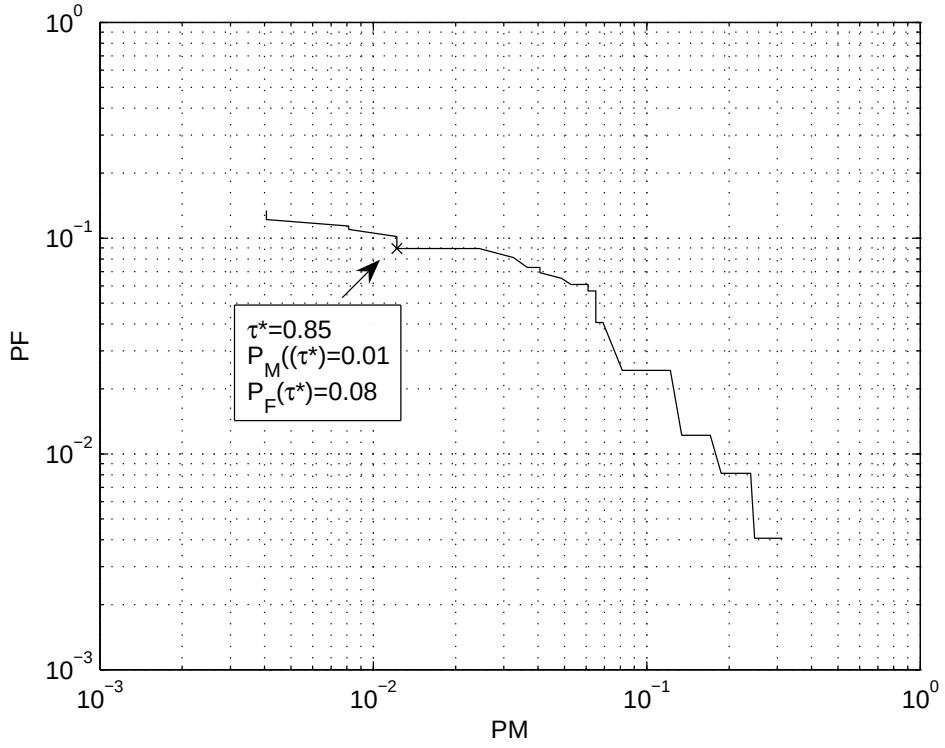


Figure 5.4. Empirical ROC curve for one-dimensional approach

Then, empirical estimated total probability of error $\left(\hat{\mathrm{P}}_{e,emp}(\tau)\right)$ can be calculated via the equation 4.9 The empirical ROC curve for empirical distributions over different

value of τ is given in Figure 5.4. According to the Figure, the minimum value of the empirical estimated total probability of error $\left(\hat{\mathrm{P}}_{e,emp}(\tau^*)\right)$ is found to be 0.045 where $\hat{\mathrm{P}}_{M,emp} = 0.01$ and $\hat{\mathrm{P}}_{F,emp} = 0.08$ for $\tau^* = 0.85$ under equal priors assumption.

6. ERROR BOUNDS OF THE METHODS

The probability of error (c.f. 4.5) might not be accurate due to the integrals in calculation of it. Therefore, the upper bounds for probability of error is generally used [22].

Definition 6.0.1 *The Chernoff Bound for the probability of error in binary hypothesis test is defined as:*

$$P(error) \leq P^{\beta}(H_0)P^{1-\beta}(H_1)\int p^{\beta}(\mathbf{x}|H_0)p^{1-\beta}(\mathbf{x}|H_1)d\mathbf{x} \quad for \quad 0 \leq \beta \leq 1. \tag{6.1}$$

where $P(H_0)$ and $P(H_1)$ are the prior probabilities of the hypothesis H_0 and H_1, respectively. [22]

Remark 6.0.1 *If the conditional probabilities are Gaussian with mean vectors $\bar{\mu}_j$ and covariance matrices $\mathbf{\Sigma}_j$ for $j = \{0, 1\}$, the integral in equation 6.1 can be calculated as;*

$$\int p^{\beta}(\mathbf{x}|H_0)p^{1-\beta}(\mathbf{x}|H_1)d\mathbf{x} = e^{-k(\beta)}$$

where

$$k(\beta) = \frac{\beta(1-\beta)}{2}(\bar{\mu}_2 - \bar{\mu}_1)^T[\beta\mathbf{\Sigma}_1 + (1-\beta)\mathbf{\Sigma}_2]^{-1}(\bar{\mu}_2 - \bar{\mu}_1) + \frac{1}{2}\log\frac{\beta\mathbf{\Sigma}_1 + (1-\beta)\mathbf{\Sigma}_2}{|\mathbf{\Sigma}_1|^{\beta}|\mathbf{\Sigma}_2|^{1-\beta}}$$

[22]

Since the conditional probabilities of multi-dimensional data are assumed to be Gaussian ($\mathrm{p}(\mathbf{x}|H_j) \sim \mathcal{N}(\bar{\mu}_j, \mathbf{\Sigma}_j)$), the equation for the Chernoff bound for multi-dimensional

approach can be given by

$$\begin{aligned}
\mathrm{P}(\text{error}) &\leq \mathrm{P}^{\beta}(H_0)\mathrm{P}^{1-\beta}(H_1)\int \mathrm{p}^{\beta}(\mathbf{x}|H_0)\mathrm{p}^{1-\beta}(\mathbf{x}|H_1)d\mathbf{x} \\
&= \mathrm{P}^{\beta}(H_0)\mathrm{P}^{1-\beta}(H_1)\exp\{-\frac{\beta(1-\beta)}{2}(\bar{\mu}_1-\bar{\mu}_0)^T[\beta\mathbf{\Sigma}_0+(1-\beta)\mathbf{\Sigma}_1]^{-1}(\bar{\mu}_1-\bar{\mu}_0) \\
&- \frac{1}{2}\log\frac{|\beta\mathbf{\Sigma}_0+(1-\beta)\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|^{\beta}|\mathbf{\Sigma}_1|^{1-\beta}}\}
\end{aligned}$$

Having the estimates of mean vectors $\bar{\mu}_j$ and covariance matrices $\mathbf{\Sigma}_j$ in equations 4.2, 4.3 and under equal prior assumption ($\mathrm{P}(H_0)=\mathrm{P}(H_1)=0.5$) the estimated $\hat{\mathrm{P}}(\text{error})$ can be written as

$$\begin{aligned}
\hat{\mathrm{P}}(\text{error}) &= 0.5\exp\{-\frac{\beta(1-\beta)}{2}(\hat{\bar{\mu}}_1-\hat{\bar{\mu}}_0)^T[\beta\hat{\mathbf{\Sigma}}_0+(1-\beta)\hat{\mathbf{\Sigma}}_1]^{-1}(\hat{\bar{\mu}}_1-\hat{\bar{\mu}}_0) \\
&- \frac{1}{2}\log\frac{|\beta\hat{\mathbf{\Sigma}}_0+(1-\beta)\hat{\mathbf{\Sigma}}_1|}{|\widehat{\mathbf{\Sigma}}_0|^{\beta}|\hat{\mathbf{\Sigma}}_1|^{1-\beta}}\}
\end{aligned}$$

Next, the equation for the Chernoff bound under Gaussianity assumption of one-dimensional data $\big(\mathrm{p}(y|H_j)\sim\mathcal{N}(\mu_j,\sigma_j^2)\big)$ for one-dimensional approach is written by

$$\begin{aligned}
\mathrm{P}(\text{error}) &\leq \mathrm{P}^{\beta}(H_0)\mathrm{P}^{1-\beta}(H_1)\int \mathrm{p}^{\beta}(y|H_0)\mathrm{p}^{1-\beta}(y|H_1)dy \\
&= \mathrm{P}^{\beta}(H_0)\mathrm{P}^{1-\beta}(H_1)\exp\left\{-\frac{\beta(1-\beta)}{2}\frac{(\mu_1-\mu_0)^2}{\beta\sigma_0^2+(1-\beta)\sigma_1^2}-\frac{1}{2}\log\frac{|\beta\sigma_0^2+(1-\beta)\sigma_1^2|}{|\sigma_0^2|^{1-\beta}|\sigma_1^2|^{1-\beta}}\right\}
\end{aligned}$$

Again, having estimates of means μ_j and variances σ_j^2 from equations 5.3, 5.4 and under equal prior assumption the estimated $\hat{\mathrm{P}}(\text{error})$ can be written as

$$\hat{\mathrm{P}}(\text{error})=0.5\exp\left\{-\frac{\beta(1-\beta)}{2}\frac{(\hat{\mu}_1-\hat{\mu}_0)^2}{\beta\hat{\sigma}_0^2+(1-\beta)\hat{\sigma}_1^2}-\frac{1}{2}\log\frac{|\beta\hat{\sigma}_0^2+(1-\beta)\hat{\sigma}_1^2|}{|\hat{\sigma}_0^2|^{1-\beta}|\hat{\sigma}_1^2|^{1-\beta}}\right\}$$

The estimated upper bounds for the probability of errors are given in the Figure 6.1. The Chernoff bound for optimal τ from the Figure 6.1 is found as 0.09 and 0.10 for multi-dimensional and one-dimensional approaches, respectively. As it can be seen from the Figure 6.1, the Chernoff bound is looser for extreme values of β and tighter for intermediate values. Then Bhattacharyya bound is defined for $\beta=\frac{1}{2}$ for simplicity.

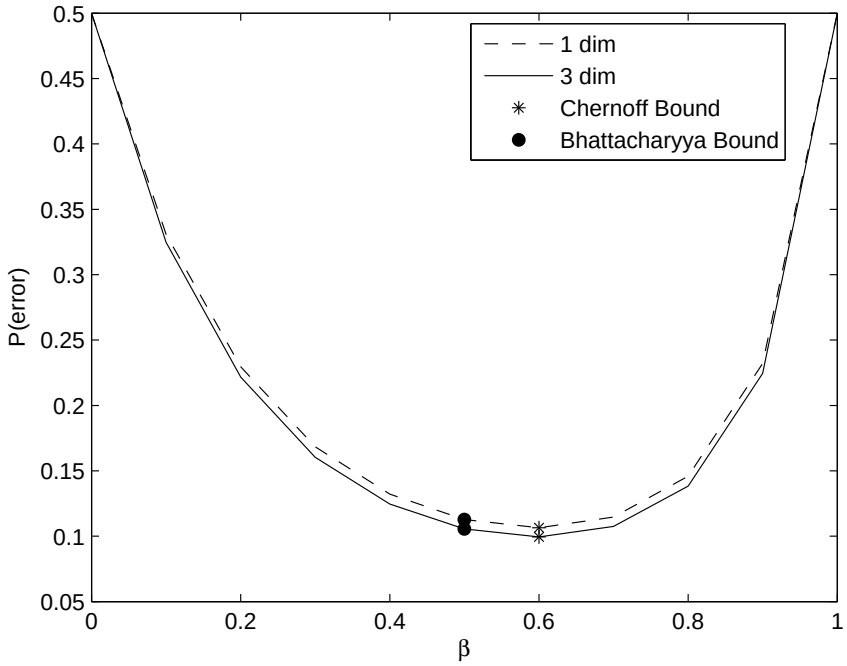


Figure 6.1. Comparison of upper bound of errors for two approaches

Bhattacharyya Bounds are found as 0.11 and 0.10 (seen from Figure 6.1) i.e. the probability of errors of the methods can not be larger than 0.11 and 0.10 for one-dimensional and multi-dimensional approaches, respectively.

7. DISCUSSION

This study's hypothesis states that the wheeze and non-wheeze parts of the lung sound signals can be detected efficiently by using their distinct behaviors in time-frequency domain. For this purpose, three different features - Renyi Entropy, f_{50}/f_{90} ratio and mean crossing irregularity- are defined and extracted before proposing the methods. Those features are shown to be significantly different for two classes in Chapter 3. So that, each signal from wheeze and non-wheeze classes can be defined in three-dimensional space.

Next, two different (although they are similar at some points) detection rules are proposed. First approach is multi-dimensional approach which uses the data in three-dimensional space and applies detection rule under Gaussianity assumption of three-dimensional data. Second approach is one-dimensional approach. It applies detection rule under Gaussianity assumption of one-dimensional data which is the projected version of three dimensional data onto one-dimensional space. Here, the three-dimensional data is projected onto a line where the two classes, wheeze and non-wheeze, are separated best by using FDA. Hence, the one-dimensional approach differs from the multi-dimensional approach in dimensionality reduction.

Gaussian assumption of both three-dimensional data and one-dimensional data is verified via kurtosis and KL distance. ML and unbiased estimates are used for the parameters of Gaussian distribution. Then, detection rules are proposed for these estimated parameters. The performance of the approaches are tested considering the probability of error both theoretically and experimentally.

The minimum value of the estimated probability of error, i.e. theoretical error is found as 0.04 for both approaches. Experimental error is calculated over the data set. Considering the false positives and false negatives among the existent data of 245 windows from each classes, the minimum value of empirical estimated probability of error, i.e. experimental error, is found to be 0.045 for both approaches. Note that,

although the total probability of error are the same, the probability of miss and false alarm values are different in two methods.

The estimated probability of errors might not be their exact value due to might be some possible miscalculations in the integral during the theoretical error calculation. Furthermore, the limited data may not represent the accurate value of experimental probability of error. Therefore upper-bounds of the probability of error are utilized. One of the well-known Chernoff bound is found for both approaches. The Chernoff bounds for probability of error is found as 0.10 and 0.09 for multi and one-dimensional approaches. Since, finding the Chernoff bound might be difficult in some cases, Bhattacharyya bound which is looser bound than Chernoff bound but easier to compute is also calculated. The Bhattacharyya bounds are 0.11 and 0.10 for one and multi-dimensional approaches, respectively. Hence, the probability of error could not be larger than these bounds.

Recall that the detection rules are applied after Gaussian fitting in two approaches. The difference is the dimensionality reduction of feature space to one-dimensional space in one-dimensional approach. The results of two approaches are almost the same which might stem from two different reasons. Firstly, when the three-dimensional data is projected onto one-dimensional space where the data can be separated best, there might not be any information loss. In other words, one of the feature among three features might be almost sufficient to separate two classes. Secondly, the decision rules for both classes have quadratic and linear terms which might result with the similar probability of errors.

On the other hand one-dimensional approach can still be said to be more efficient than the multi-dimensional approach because of the reduction of the estimated parameters in Gaussian fitting step. In multi dimensional case 9 parameters are required to be estimated (6 for the covariance matrix and 3 for the mean vector) for each distribution. However, there are only 2 parameters required to be estimated for each distribution (1 for mean and 1 for variance). Moreover, for one-dimensional approach, calculation process is shorter than that of the multi-dimensional approach.

In conclusion, these empirical and theoretical results and upper bounds confirm that the proposed methods in both dimensions are successful in the detection of wheeze. Hence, it can be concluded that the hypothesis of this thesis is supported. Furthermore, the performance of the approaches can be improved, potentially, either by adding new features and then defining the data in larger space or utilizing new techniques for dimensionality reduction. Moreover, these methods will be employed in developing algorithms for online wheeze detection in a given lung sound. These results contribute tot he efforts of computerized lung sound analysis in respiratory disorders. The approaches differs from previous studies since a combination of detection and estimation theory, pattern recognition and information theory are applied. The concepts used in this study may be applied to other adventitious sounds with an aim to make lung auscultation a more objective and valuable diagnosis method.

8. CONCLUSION

In this study, we proposed methods for the detection of wheeze in respiratory sound signals. Wheeze is the abnormal lung sound which is an indicator of asthma and COPD diseases. Therefore, efficient detection algorithm might help the diagnosis of obstructive diseases. Up to now, although there have been some previous works on the detection of wheeze, there is not any standard algorithm to detect wheeze in respiratory sounds. The lack of world-wide common shared respiratory data adds to the difficulty of comparison.

This study not only proposes two methods but also defines new features to be used in the classification and detection of wheeze. The features and methods are verified to be successful in detection of wheeze. The performance rates of the methods over the 245 samples from each class is found as %95.5 in both multi and one-dimensional approaches.

The study can be summarized basically in six steps. First, the data is acquired from asthma patients by using the lung-analysis system with 14 channels [18]. Then, the windows of wheeze and non-wheeze signals are selected by an expert. Based on the differences of wheeze and non-wheeze, three different features are defined. Then, Gaussian fitting is applied in three-dimensional space for multi-dimensional approach and one-dimensional space for one-dimensional approach. Next, detection rule is applied for these estimated Gaussian pdf's. Finally, the empirical and theoretical performance of the methods are computed and compared.

The proposed approaches can be considered as novel in this area. The previous studies generally utilized the concepts from artificial neural network while this study uses the combination of detection and estimation theory, pattern recognition and information theory. The results of the methods confirm the reliability and efficiency of the proposed methods.

As mentioned in chapter 7, this study might broaden the horizons of the researchers in this area by enabling them to consider the problem of detection of wheeze from different points of view. Furthermore, the proposed methods may be utilized and improved for the detection and classification of other abnormal sounds.

APPENDIX A: WELCH METHOD for PSD ESTIMATION

The Welch method is obtained by averaging the windowed periodograms over overlapping data segments. Mathematically, let

$$y_j(t) = y((j-1)K + t), \quad \text{for} \quad t = 1, ..., M \quad \text{and} \quad j = 1, ..., S$$

denote the j^{th} data segment where $(j-1)K$ is the starting point for the j^{th} sequence of observations. The value recommended for K in the Welch method is $K = M/2$, i.e. 50% between successive data segments. The windowed periodogram corresponding to $y_j(t)$ is computed as

$$\hat{\phi}_j(w) = \frac{1}{2}\left|\sum_{t=1}^{M} v(t)y_j(t)e^{-iwt}\right|^2 \tag{A.1}$$

where P denotes the "power" of the temporal window $\{v(t)\}$:

$$P = \frac{1}{M}\sum_{t=1}^{M}|v(t)|^2$$

The Welch method is found by averaging the windowed periodograms in equation A.1:

$$\hat{\phi}_W(w) = \frac{1}{S}\sum_{j=1}^{S}\hat{\phi}_j(w)$$

[19]

APPENDIX B: Q Function

The Q function is often used to evaluate the probability of error analysis in detection problems. Let Z be a Gaussian random variable with mean 0 and variance 1, then $Q(.)$ function can be defined as

$$Q(x) = P(Z > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

If W be a Gaussian random variable with mean μ and variance σ^2, then the result

$$\begin{aligned} P(W > x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(w-\mu)^2}{2\sigma^2}} dw \\ &= \int_{\frac{x-\mu}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= Q(\frac{x-\mu}{\sigma}) \end{aligned}$$

will be found by change of variables $y \triangleq \frac{w-\mu}{\sigma}$ [23].

REFERENCES

1. Gavriely N., *Breath Sounds Methodology*, p. 162, CRC Press, Boca Raton, FL, 1995.

2. Pasterkamp H., S. S. Kraman, and G. R. Wodicka, "Advances beyond the stethoscope", *Am J Respir Crit Care Med*, Vol. 156, pp. 974-87, September 1997.

3. Vanuccini L., J. E. Earis, P. Helistö, B. M. G. Cheetham, M. Rossi, A. R. A. Sovijarvi and J. Vanderschoot, "Capturing and Preprocessing of Respiratory Sounds", *European Respiratory Review*, Vol. 10, No. 77, pp. 616-620, 2000.

4. Earis J. E. and B. M. G. Cheetham, "Current Methods Used for Computerized Respiratory Sound Analysis", *European Respiratory Review*, Vol. 10, No. 77, pp. 586-590, 2000.

5. Sovijärvi A. R. A., L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi and J. E. Earis, "Characteristics of Breath Sounds and Adventitious Respiratory Sounds" ,*European Respiratory Review*, Vol. 10, No. 77, pp. 591-596, 2000.

6. Gavriely N., M. Nissan, A. H. Rubin and D. W. Cugell., "Spectral characteristics of chest wall breath sounds in normal subjects", *Thorax*, Vol. 50, No: 12, pp. 1292-1300, 1995.

7. Sovijarvi A. R. A., F. Dalmasso, J. Vanderschoot, L. P. Malmberg, G. Righini and S. A. T. Stoneman, "Definition of Terms for Respiratory Sounds", *European Respiratory Review*, Vol. 10, No. 77, pp. 597-610, 2000.

8. Charbonneau G., E. Ademovic, B. M. G. Cheetman, L. P. Malmberg, J. Vanderschoot and A. R. A. Sovijarvi, "Basic Techniques for Respiratory Sound Analysis", *European Respiratory Review*, Vol. 10, No. 77, pp. 626-635, 2000.

9. Forkheim K. E., D. Scuse and H. Pasterkamp, "A comparison of neural network models for wheeze detection", *WESCANEX 95. Communications, Power, and Computing. Conference Proceedings. IEEE* , Vol.1, pp.214-219 vol.1, 15-16 May 1995

10. Kandaswamy A. , C. Kumar, R. Ramanathan, S. Jayaraman and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients",em Computers in Biology and Medicine, Vol. 34 , no. 6 , pp. 523 - 537, 2004.

11. Rietveld S., M. Oud and E. H. Dooijes, "Classification of Asthmatic Breath Sounds: Preliminary Results of the Classifying Capacity of Human Examiners versus Artificial Neural Networks", *Computers and Biomedical Research*, Vol. 32, No. 5, pp. 440-448, October 1999.

12. Dokur Z., "Respiratory sound classification by using an incremental supervised neural network",*Pattern Anal Applic*, 2008, ISSN 1433-755X.

13. Güler İ., H. Polat and U. Ergün, "Combining Neural Network and Genetic Algorithm for Prediction of Lung Sounds", *Journal of Medical Systems*, Vol. 29, No. 3, June 2005.

14. Taplidou S. A., L. J. Hadjileontiadis, T. Penzel, V. Gross and S. M. Panas, "WED: An efficient wheezing-episode detector based on breath sounds spectrogram analysis",*Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE*, Vol.3, pp. 2531-2534, 17-21 Sept. 2003.

15. Taplidou S. A., L. J. Hadjileontiadis, I. K. Kitsas, K. I. Panoulas, T. Penzel, V. Gross and S. M. Panas, "On Applying Continuous Wavelet Transform in Wheeze Analysis"',*Engineering in Medicine and Biology Society, 26th Annual International Conference of the IEEE*, Vol.2, pp. 3832-3835, 1-5 Sept. 2004.

16. Taplidou S. A. and L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds", *Computers in Biology and Medicine* , Vol.

37, No. 8, pp. 1073 - 1083, 2007.

17. Robert L. W. , J. E. Hodgkin and B. Lopez, *Fundamentals of Lung and Heart Sounds*, Mosby, Third Edition, 2004.

18. İ. Şen, "A Multi-channel Device for Respiratory Sound Data Acquisition and Transient Detection", *M.S. Thesis*, Boğaziçi University, 2005.

19. Petre S. and R. Moses *Spectral Analysis of Signals*, PEARSON-Prentice Hall, 2005.

20. Henry S. and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, Prentice Hall, Third Edition, 2002.

21. Thomas M. C. and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience Publication, 1991.

22. Richard O. D. , P. E. Hart and D. G. Stork, *Pattern Classification*, Wiley-Interscience Publication, Second Edition, 2001.

23. Todd K. M. and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, 2000.

24. Poor H. V., *An Introduction to Signal Detection and Estimation*, Springer–Verlag, New York, 1988.