TIME-FREQUENCY FEATURE ANALYSIS

by

Behnaz Ghoraani, B.Sc, M.Sc.,
Sharif University of Technology, Tehran, Iran, 1998,
Amir Kabir University of Technology, Tehran, Iran, 2000,

A dissertation
presented to Ryerson University
in partial fulfillment of the
requirement for the degree of
Doctor of Philosophy
in the Program of
Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2010

Author's Declaration

I hereby declare that I am the sole author of this thesis.
I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.
Signature

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

Abstract

Time-frequency Feature Analysis

© Behnaz Ghoraani, 2010

Doctor of Philosophy

in the Program of

Electrical and Computer Engineering,

Ryerson University.

Most of the real-world signals in nature are non-stationary, i.e., their statistics are time variant. Extracting the time-varying frequency characteristics of a signal is very important in understanding the signal better, which could be of immense use in various applications such as pattern recognition and automated-decision making systems. In order to extract meaningful time-frequency (TF) features, a joint TF analysis is required. The proposed work is an attempt to develop a generalized TF analysis methodology that exploits the benefits of TF distribution (TFD) in pattern classification systems as related to discriminant feature detection and classification.

Our objective is to introduce a unique and efficient way of performing non-stationary signal analysis using adaptive and discriminant TF techniques. To fulfill this objective, in the first point, we build a novel TF matrix (TFM) decomposition that increases the effectiveness of segmentation in real-world signals. Instantaneous and unique features are extracted from each segment such that they successfully represent joint TF structure of the signal.

In the second point, based on the above technique, two unique and novel discriminant TF analysis methods are proposed to perform an improved and discriminant feature selection of any non-stationary signals. The first approach is a new machine learning method that identifies the clusters of the discriminant features to compute the presence of the discriminative pattern in any given signal, and classify them accordingly. The second approach is a discriminant TFM (DTFM) framework, which is a combination of TFM decomposition and discriminant clustering techniques. The developed DTFM analysis automatically identifies the differences between different classes as the distinguishing structure, and uses the identified structure to accurately classify and locate the discriminant structure in the signal.

The theoretical properties of the proposed approaches pertaining to pattern recognition and detection are examined in this dissertation. The extracted TF features provide strong and successful characterization and classification of real and synthetic non-stationary signals. The proposed TF techniques facilitate the adaptation of TF quantification to any feature detection technique in automating the identification process of discriminatory TF features, and can find applications in many different fields including biomedical and multimedia signal processing.

Acknowledgments

I am heartily thankful to my supervisor, Professor Sri Krishnan, for his wonderful guidance throughout all the stages of my PhD studies. Without his continued encouragement and support, this dissertation would not have been possible. I am also thankful for the excellent example Professor Sri Krishnan has provided as a successful researcher, wonderful teacher and role model.

I wish to express my sincere gratitude to Dr. Karthi Umapathy whose help and technical discussions have been invaluable to me. I offer my regards and blessings to my colleagues and staff at Signal Analysis Laboratory (SAR) and Ryerson University who supported me in every aspect during the completion of my studies.

I would like to thank Dr. Vijay Chauhan for providing the exceptional opportunity for me to collaborate with his Electrophysiology team at Toronto General Hospital. I am grateful to Dr. Raja Selvaraj for the enlightening and informative discussions we had about the T wave alternans. I would like to acknowledge Adrian Suszko and Dhinesh Sivananthan for their wonderful assistance in providing the T wave alternans database.

Last but not least, I would like to gratefully acknowledge the financial support of Canada Research Chairs program, NSERC, OGS and OGSST, University Health Network for the clinical input, and the LastWave software group for providing the software for signal decomposition.

Dedication

To my wonderful husband and son, Hamid and Kiarash, for their unconditional love, enormous support and continuous encouragement...

Contents

List of Tables

List of Figures

Table of Abbreviations

Abbr.	Full Name	Abbr.	Full Name
ANR	Alternans-to-noise ratio	MMA	Modified moving average method
BCH	Bose-Chaudhuri-Hocquenghem	MP	Matching pursuit
BER	Bit error rate	MSE	mean squared error
CD	Complex demodulation	NMF	Non-negative matrix factorization
ChSR	Chirp-to-signal ratio	NMFDB	NMF Discriminant base
CM	Correlation method	PCA	principal component analysis
DCT	Discrete cosine transform	pdf	Probability density function
DNMF	Discriminative NMF	PTFD	positive TFD
DPPT	Discrete polynomial phase transform	REP coding	Repetition coding
DTF	discriminant TF	RMS	Root mean square
DTFM	Discriminant TFM	ROD	Region of discrimination
EEG	Electroencephalogram	SCD	Sudden cardiac death
FEC	Forward error correction	SM	Spectral method
HRT	Hough-Radon Transform	SNR	Signal-to-noise ratio
HVS	human visual system	SOTM	Self organizing tree map
ICA	Independent component analysis	STFT	Short-time Fourier transform
JND	just noticeable difference	SVD	Singular value decomposition
LDA	Linear discriminant analysis	TFD	Time-frequency distribution
MCE	Minimum cross entropy	TFM	Time-frequency matrix
MD	Matrix decomposition	TWA	T wave alternans (TWA)
MFCC	Mel-frequency cepstral coefficient		

List of Symbols

Symbol	Meaning
$\exp(x)$	Exponential of x.
π	3.14
$\int_a^b x(t)dt$	Integral of function $x(t)$, when t is changing from a to b.
$x^*(t)$	Conjugation of $x(t)$.
$\|x(t)\|$	Absolute value of $x(t)$.
$\text{Min}(x(t))$	Minimum of $x(t)$.
$\text{Max}(x(t))$	Maximum of $x(t)$.
$\sum_{t=a}^{b} x(t)$	Summation of discrete function $x(t)$ when t is changing from a to b.
$\delta(t-t_0)$	Dirac delta function at time t_0.
$\sin(2\pi f_0 t)$	Sinusoidal wave with frequency of f_0.
x^2	Square of x.
$\sqrt{(x(t))}$	Square root of $x(t)$.
$\mathcal{FFT}_x(f)$	Fourier transform of $x(t)$.
$\text{Real}\{x(t)\}$	Real part of function $x(t)$.
$\text{Log}(x(t))$	The logarithm of $x(t)$ to base of 10.
$\prod_{t=a}^{b}(x(t))$	Product of $x(t)$ samples when t is changing from a to b.
$m!$	Factorial of m.
$\binom{m}{n}$	Combination function.
$\mathbf{H}^{-1}$	Inverse of matrix **H**.
$\mathbf{H}^T$	Transpose of matrix **H**.
$\langle \mathbf{A}, \mathbf{B} \rangle$	The inner product between two matrices, **A** and **B**.
$\nabla^P f_{\mathbf{H}}(\mathbf{W})$	The projected gradient of function f with respect to **W** and constant **H**.
$D(A,B)$	The square error between A and B.

Chapter 1

INTRODUCTION

1.1 Motivation

OVER the last century, a significant attention has been paid to the field of signal analysis. Signals provide valuable information about the physical world. Human beings have learned over the years how to use this information to find existing or appearing trends in the surroundings, and use the information toward their own benefit. For example, looking through the weather temperature signal over the last several years, scientists realized that the global warming was occurring and encouraged humans to initiate preventive actions against it. Or, studying the electrocardiogram (ECG) of a patient's heart, physicians find out critical information about the patient's health that is used in developing new diagnosis and treatment tools.

The advancement in sensor technology made it possible to gather huge amounts of data, which on the one hand, extends the applicability of signal analysis to a wide variety of fields, such as, communication, security, biomedicine, biology, physics, finance and geology. But on the other hand, this huge data makes demands for advanced signal analysis techniques to effectively process the gathered data. Before the recent significant growth in technology, most signal analysis techniques were performed by humans. However, currently due to the substantial progression in both data collection and processor technologies, there is a great interest in advanced signal analysis. We can design algorithms, develop models, and make informed decisions based on the models.

Fig. 1.1 summarizes the advantages that advanced signal analyses offer over the simple manual method. Unlike manual data processing, advanced methods are reliable, objective, and more

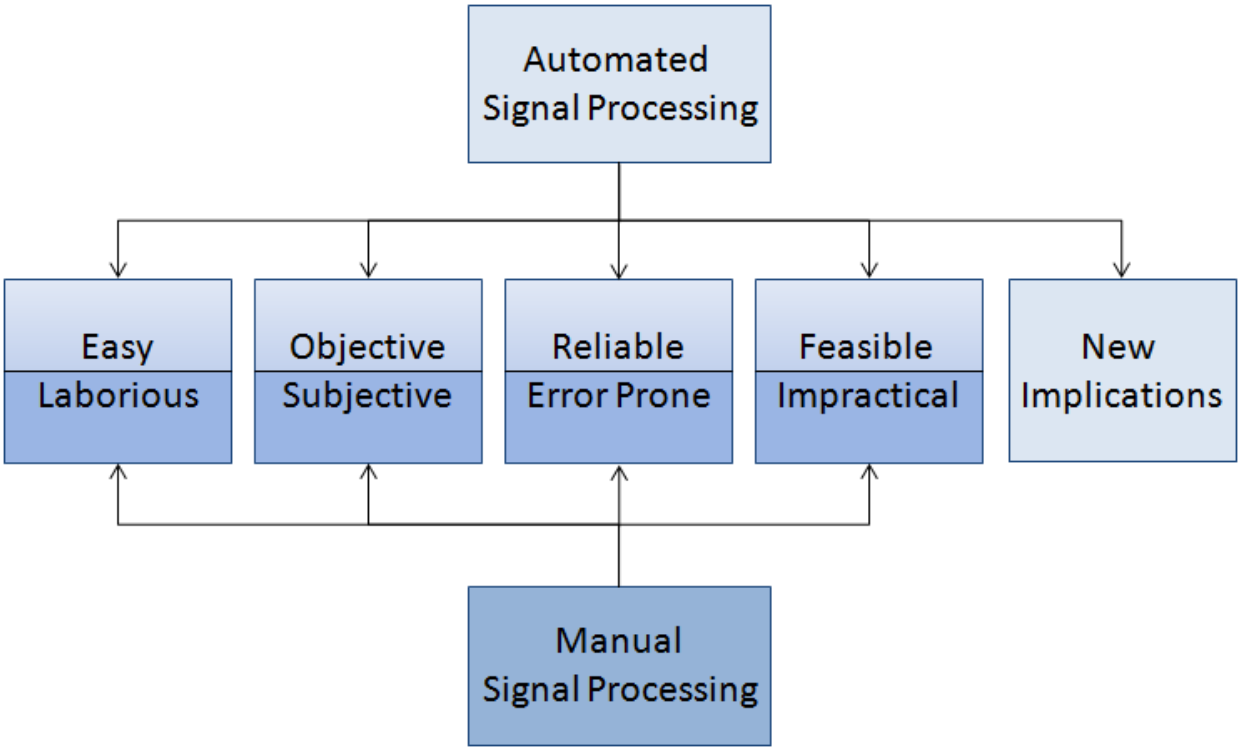


Figure 1.1: Advantages of advanced signal analysis techniques over manual method.

efficient. Some highlighted differences between these two approaches are as follows:

- Usually, the information of interest resides on a small portion of data, and as a result, finding it within large amount of the data could be a laborious task. For instance, in an electroencephalogram (EEG) signal, the data are collected for several hours from the human brain, but in epilepsy and other brain disorder screening applications, the signals of interest are bursts with durations of as short as 100ms. The brief presence of the abnormality bursts over long hours of EEG signals creates difficulties in finding them and as a result diagnosing the problem.

- In addition, manual data screening is subjective; i.e. depending on the level of expertise, different results can be reported from the same data. However, if the right model is learned in the automated data processing, it assists specialists to obtain an objective outcome.

- Furthermore, the manual data processing could be error-prone as humans tend to perform a

task more accurately in the beginning, and more carelessly later on during the day. Also, the link between the brain function to feelings limits people's ability to perform consistently and accurately while completing a given task. However, automated methods are not restricted by such limitations.

- Additionally, performing some data processing tasks is possible only with the use of automatic signal analysis techniques. An example of such an application is in hearing aids where determining the environment enables us to build better hearing aid instruments with automatic switching features that change adaptively according to the environment. This application is feasible only with an automatic system.

- Last but not least, automated signal processing systems make it possible to develop new quantities that might not be captured by human perception. One example of such quantities is T wave alternans (TWA), also called repolarization alternans, which is emerging as an important prognostic marker for sudden cardiac death in patients with heart disease. TWA is invisible, and can be only measured with the aid of an automated signal processing method.

Having mentioned the importance of advanced data processing techniques, the main focus of this dissertation is to develop an improved and highly effective signal processing techniques as related to biomedical and multimedia applications, which are considered to be among the most dynamic and challenging subjects in the science of signal processing.

1.2 What Is a Signal?

A signals can be considered as a unique communication that occurs between the physical environment and a human being. This communication can be either voluntary or stimulated. The former indicates the cases where the signal is measured directly from a physical quantity. An example is the measurement of temperature signals, or recording ECG signals of a patient. In the latter case, a signal is used to stimulate the communication with the physical world. For instance, ultrasound waves, which are acoustic signals, are beamed into the body causing return echoes that are

recorded to produce pictures of fetuses in the human womb; or in radar, electromagnetic signals are used to identify the range, altitude, direction, or speed of aircrafts or weather formations.

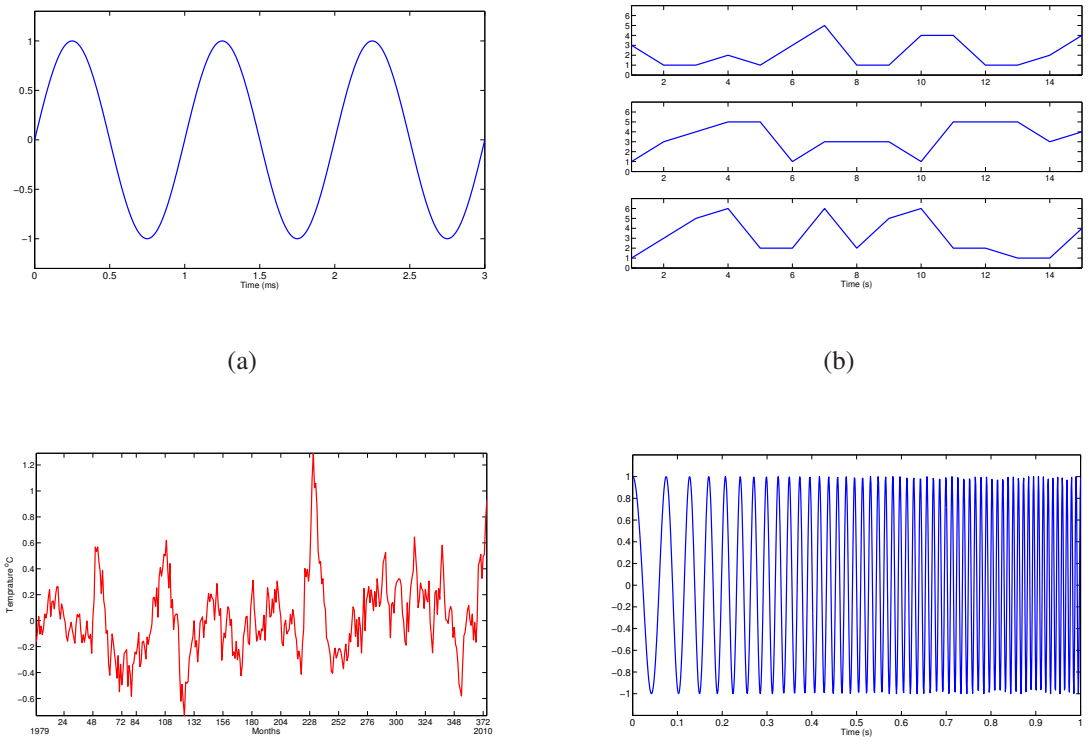
1.2.1 Signal Categories

There are many diverse types of signals ranging from simple structures to very complex ones. A sinusoidal signal is the simplest example of a signal. The sinusoidal signal shown in Fig. 1.2(a) has a frequency of 1 kHz and amplitude of one unit. The value of the signal at each time sample can be predicted from the equation of $y = \sin(2\pi 1000t)$, where y represents the value of the signal at time t. This figure shows three cycles of this sinusoid. However, even if the plot is continued, there will be no unpredictability in the signal value at any given time sample. The mathematical term for such signals is deterministic. The opposite term of deterministic is non-deterministic. An example of a non-deterministic signal is the one constructed by recording the integers seen when rolling a die at every second. Fig. 1.2(b) shows three signals constructed by throwing a die as explained above. As shown in this figure, signal values at each time are not predictable, and unlike the sinusoidal signal, the die signal cannot be described in a closed-form analytical equation.

There are two types of non-deterministic signals. An example of the first type is the die generated signal shown in Fig. 1.2(b). If each face of the die is equally likely to occur, the probability of the signal having amplitude of 1 at any time is $\frac{1}{6}$. These types of non-deterministic signals which can be expressed using their probabilistic or statistical values are called stationary signals. Now let us consider a weather temperature signal. Fig. 1.2(c) displays the global temperature signal over the last 32 years [1]. One may find a pattern for this signal depending on the season when the temperature was recorded, but in general, the temperature value at each time is unpredictable. Therefore, it is impossible to find a mathematical function, or to define an accurate or fixed probabilistic value for each of the temperature values. Such signals which are neither deterministic nor stationary are called non-stationary signal.

In summary, any signal can be classified into one of the following categories:

- **Deterministic or non-deterministic**: Deterministic signals can be usually described by analytical expressions, and are predictable for all times (past, present and future). Thus, they



(a) (b)

(c) (d)

Figure 1.2: Signal categories: (a) Deterministic - three cycles of a sinusoidal signal with frequency of 1 kHz and amplitude of one. (b) Non-deterministic and stationary - a die is rolled every second, and the integer observed is recorded as the amplitude of the signal at each second. Three signals as related to three separate trials are shown. (c) Non-deterministic and non-stationary - the global temperature signal over the last 32 years as obtained from the National Space Science and Technology Center (NSSTC) [1]. (d) Deterministic and non-stationary - a chirp signal with amplitude of one and instantaneous frequency increasing from 10 Hz to 100 Hz.

are also predictable for any arbitrary time and can be reproduced. On the other hand, those signals whose exact instantaneous values are unpredictable are called non-deterministic signals.

- **Stationary or non-stationary**: If signal statistics, such as mean and variance, are fixed over time, the signal follows a probabilistic distribution, and it therefore belongs to the stationary category. In contrast, if a signal's statistics are variable in time, the signal is denoted as non-stationary.

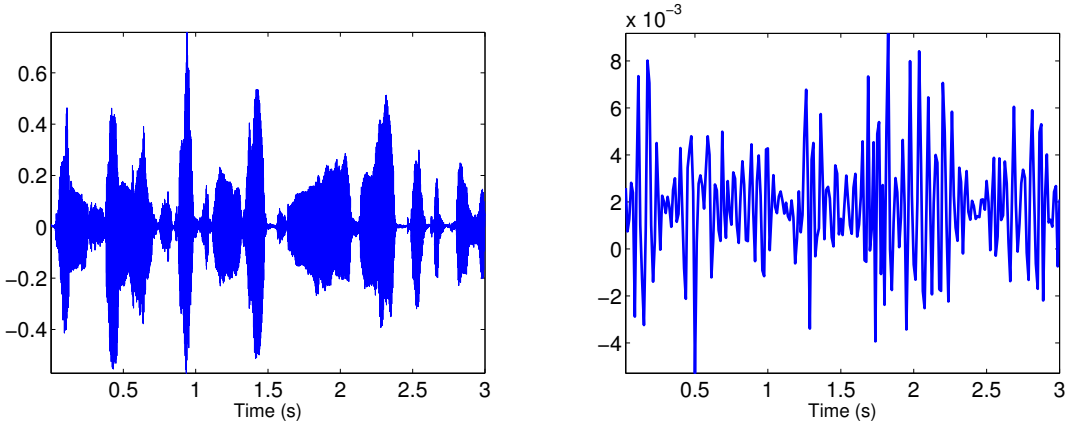
The above signal categorization leads to four types of signals as follows: 1) deterministic and stationary, e.g. a sinusoidal signal with fixed frequency and amplitude (Fig. 1.2(a)); 2) non-deterministic and stationary, e.g. signals generated in the rolling a die example (Fig. 1.2(b)); 3) deterministic and non-stationary, e.g. a linearly frequency modulated signal (Fig. 1.2(d)), which is a sinusoidal signal with frequency varying by time; and 4) non-deterministic and non-stationary, e.g. the global weather temperature signal (Fig. 1.2(c)).

1.2.2 Real-world Signals

A majority of real-world signals generated by nature (i.e., temperature or biological signals) are *non-deterministic* and *non-stationary*. For example, Fig. 1.3 shows a three second segment of a speech signal, and its first and second order statistics (mean and variance, respectively). The time-varying mean and variance indicates that the signal's statistics are varying over time, which means that the probability distribution of the signal is also time-varying. For the sake of simplicity, real-world signals are termed non-stationary signals in the rest of this dissertation, rather than non-deterministic and non-stationary.

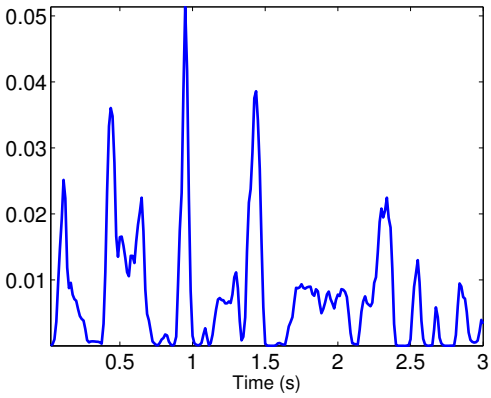
1.3 Analysis of Non-stationary Signals

The natural signals carry significant information on the generating phenomenon. Therefore, if the underlying information in these signals is analyzed properly, valuable facts could be extracted, and used to improve many aspects of life. In the previous section, real-world signals were introduced



(a) Speech signal

(b) Mean



(c) Variance

Figure 1.3: (a) A 3 s speech signal is shown as an example of real world non-stationary signal. (b) Mean of the signal is computed over moving windows of 25 ms long and 50% overlap. (c) Variance of the signal is computed over moving windows of 25 ms long and 50% overlap.

as non-stationary signals. Due to the uncertainties in these signals and their generation process, the spectral and temporal structures of real-world signals are varying over time. This property of non-stationary signals makes them the most challenging signal type. Hence, when it comes to real-world signals, in order to represent and understand the signals better, it is important to select a signal representation that reveals the varying structure in a given signal. Selection of a correct signal illustration can reflect a clear understanding about the signal production.

Considering the challenges of non-stationary signal analysis, and as the final goal to develop complex signal processing techniques to adaptively analyze real signals, we explain the available signal representation tools and their limitations.

1.3.1 Signal Representation Tools

There are three main signal representation domains as follows:

Time: The most common signal representation form is time domain which represents signal values at each time sample. Fig. 1.4(a) shows the signal of a recorded bird song. The song consists of a set of repeated short phrases starting with "WIP, WIP, WIP, WIP, WIP, WIP", and followed by a second series of repeated phrases "CHEUW, CHEUW, CHEUW, CHEW". From the time domain signal shown in Fig. 1.4(a), it can observed that the song consists of repeated short phrases with maximum and minimum of 1 and -1, respectively. Common analyses in time domain include calculation of maximum, minimum, average, and variance of the signal amplitude, energy, convolution, and correlation.

Frequency: Frequency representation of a signal represents the collected energy of the signal at each frequency. Fourier transform converts a signal from time domain to the frequency domain by projecting the signal into sinusoidal signals with frequencies varying from 0 Hz to F_s, where F_s is the sampling frequency of the signal. According to Fourier theory, the projection energy at each frequency is equal to the total energy of the signal at the corresponding frequency. In Fig. 1.4(b), the same bird song is shown in frequency domain. It can be seen from the frequency signal representation that the majority of energy occurs at the first quarter of the frequency band, spanning from 2 kHz to 4 kHz.

Joint Time-frequency (TF): In joint TF domain, a signal is represented simultaneously as instantaneous energy distribution over time and frequency. Fig. 1.4(c) shows the TF representation of the bird song. In this figure, x and y axes represent time and frequency values in the signal, and the third dimension, which is the intensity of the graph, illustrates the energy of the signal at each time and frequency. According to the intensity bar on the right side of this graph, a darker point means that the signal contains more energy at the time and frequency indicated by that point. The repeating frequency patterns that we observe over short duration of time are called frequency-modulated notes. These patterns are studied in bird vocalization and ornithology.

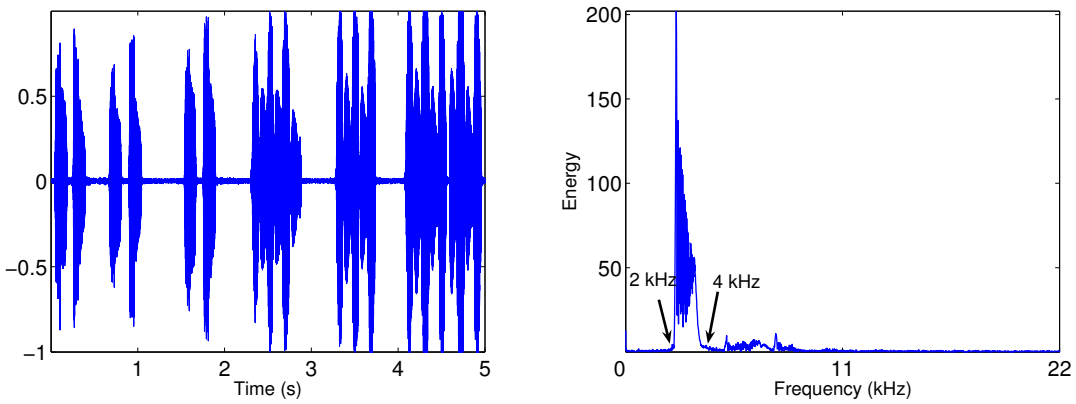
The TF plane provides all the information that can be observed in time and frequency representations. In addition, the joint TF representation relates the frequency properties of the signal to its temporal information. From the frequency representation shown in Fig. 1.4(b), we found that the energy of the bird's signal was mainly concentrated in the frequency range of 2 to 4 kHz, but the representation did not provide any specific information about the energy distribution over time. On the other hand, the time representation did not offer any information about the frequency distribution of the signal. However, the joint TF plane of the bird song successfully displayed the phrases observed in the time representation (Fig. 1.4(a)), along with the frequency values and trends at each phrase.

Depending on the application, analysis can be performed in any of the above representation domains. The purpose of this section is to identify the proper signal representation tool for real-world signals.

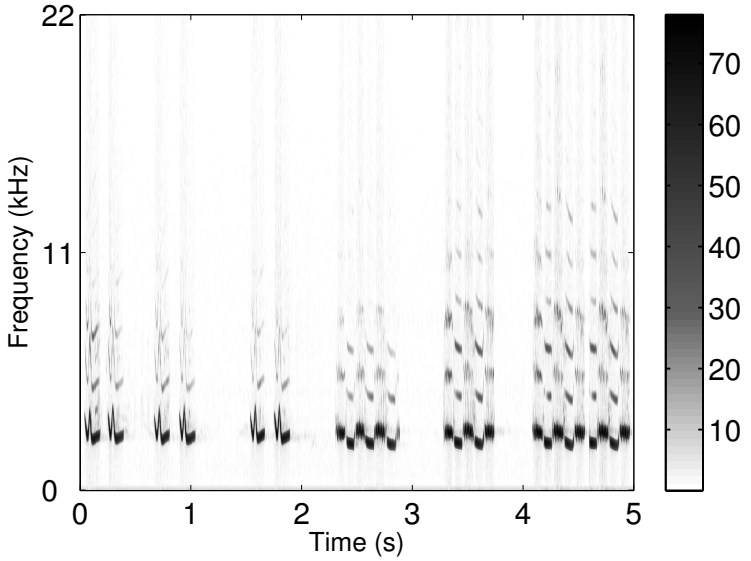
1.3.2 Selection of Signal Analysis Domain

As mentioned in the previous section, signals in most real world applications present non-stationary characteristics. The bird example in Fig. 1.4 demonstrated that the joint TF representation domain displayed more information about the structure of the bird song compared to only time or frequency domains. In this section, we create a toy example which further compares three signal representation domains as related to signal non-stationarity.

An example of a simple non-stationary signal is shown in Fig. 1.5(a). In this figure, a linearly



(a) (b)



(c)

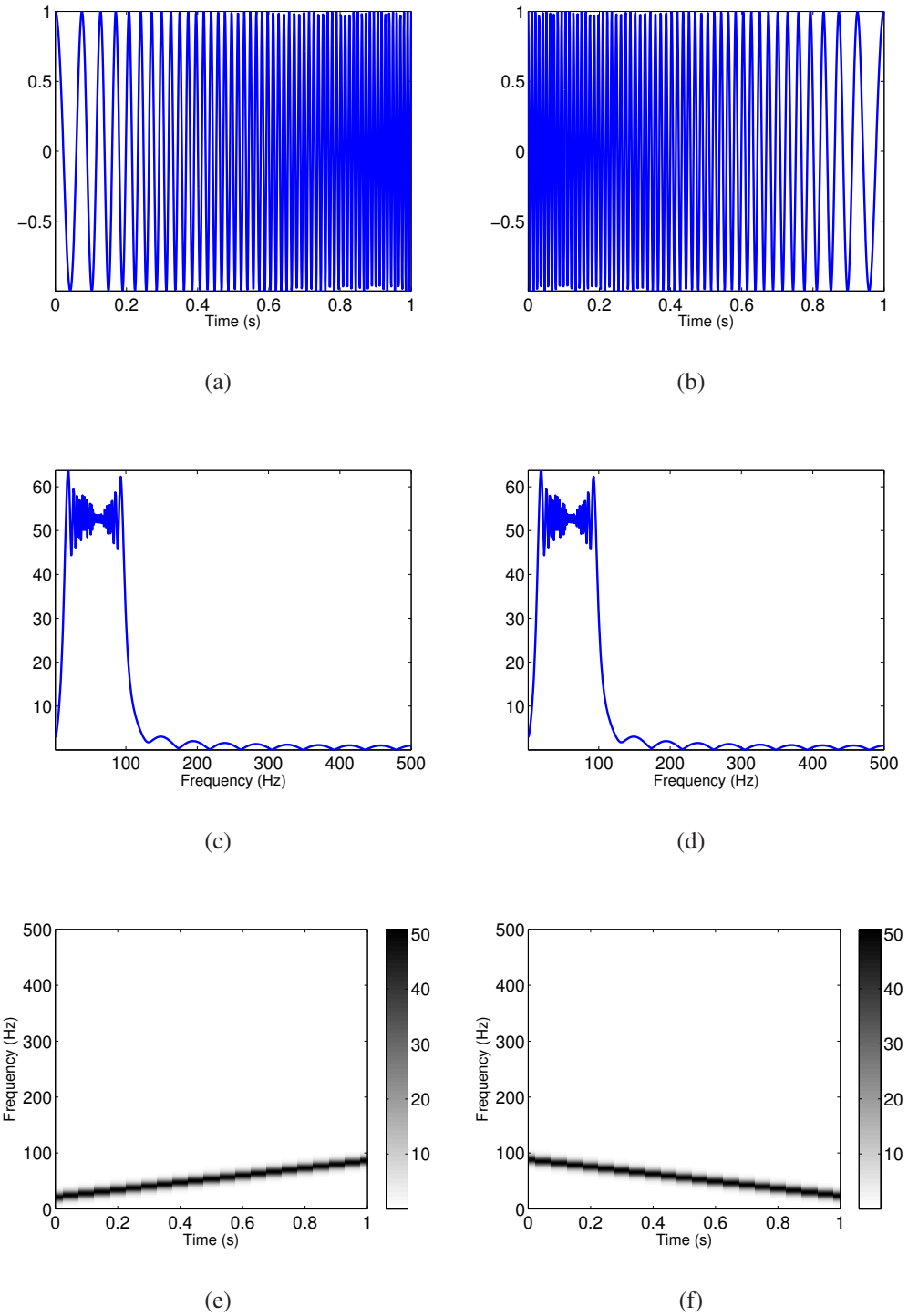
Figure 1.4: (a) 5 s of a recorded bird song at 44 kHz and 16 bits in time domain. (b) Frequency representation of the bird song is calculated using Fourier transform with 4096 point. (c) Joint TF plane of the bird song is calculate using Short Time Fourier Transform (STFT) with Kaiser window (β=5) of 256 sample long and 220 sample overlap, and 1024 point Fast Fourier Transform (FFT).

frequency modulated signal, which is called a chirp, is shown. The frequency of the chirp increases from 10 Hz to 100 Hz. Fig. 1.5(b) shows a similar chirp signal, but with frequency decreasing from 100 Hz to 10 Hz. We call the constructed increasing and decreasing chirps as $Chirp$ 1 and $Chirp$ 2. The Frequency and joint TF representations of the $Chirps$ 1 and 2 are shown in Fig. 1.5. Comparing all the three different representation domains for the two chirps, it can be seen that neither time or frequency representations are enough to represent $Chirps$ 1 and 2. The time representation shows that the frequency in $Chirp$ 1 is increasing, while it is decreasing in $Chirp$ 2. However, the frequencies of the chirps cannot be identified from the temporal signals. The frequency plots show that both chirp signals contain components in the frequency range of 10 Hz to 100 Hz, but they do not provide any information about the localization of the frequency content in time domain. The joint TF planes of both chirps are shown in Figs. 1.5(e) and 1.5(f), respectively. From the TF domain, it can be realized that the signals have the same energy over all the time samples, and the frequency of $Chirp$ 1 increases from 10 Hz to 100 Hz and vice versa in $Chirp$ 2. Unlike time or frequency representations, the TF plots can successfully track energy trend over time and frequency axes.

Joint TF signal representation provides a 3D signal domain that reveals not only temporal information, such as energy, but also frequency trend over time. As shown in bird and chirp examples, TF domain displays signal patterns that cannot be observed in the other two signal domains. Therefore, it is suggested as the most suitable signal plane for analysis of real-world signals which contain uncertainty and variability over both frequency and time.

1.3.3 Long-term Analysis vs. Short-term Analysis

Classic signal processing tools, such as Fourier analysis, are powerful tools for analyzing a stationary signal. Therefore, a short-term signal analysis has been traditionally introduced to deal with the non-stationarity of real-world signals. This approach assumes the stationarity of a signal over shorter segments, and based on this assumption, splits the signal into shorter frames through a process called signal segmentation. The short-term analysis then applies stationary tools to analyze each frame.



(a) (b) (c) (d) (e) (f)

Figure 1.5: (a) $Chirp$ 1 with sampling frequency of 1 kHz and instantaneous frequency increasing from 10 Hz to 100 Hz. (b) $Chirp$ 2 with instantaneous frequency decreasing from 100 Hz to 10 Hz. (c) Frequency representation of $Chirp$ 1. (d) Frequency representation of $Chirp$ 2. (e) Joint TF of $Chirp$ 1. (f) Joint TF of $Chirp$ 2.

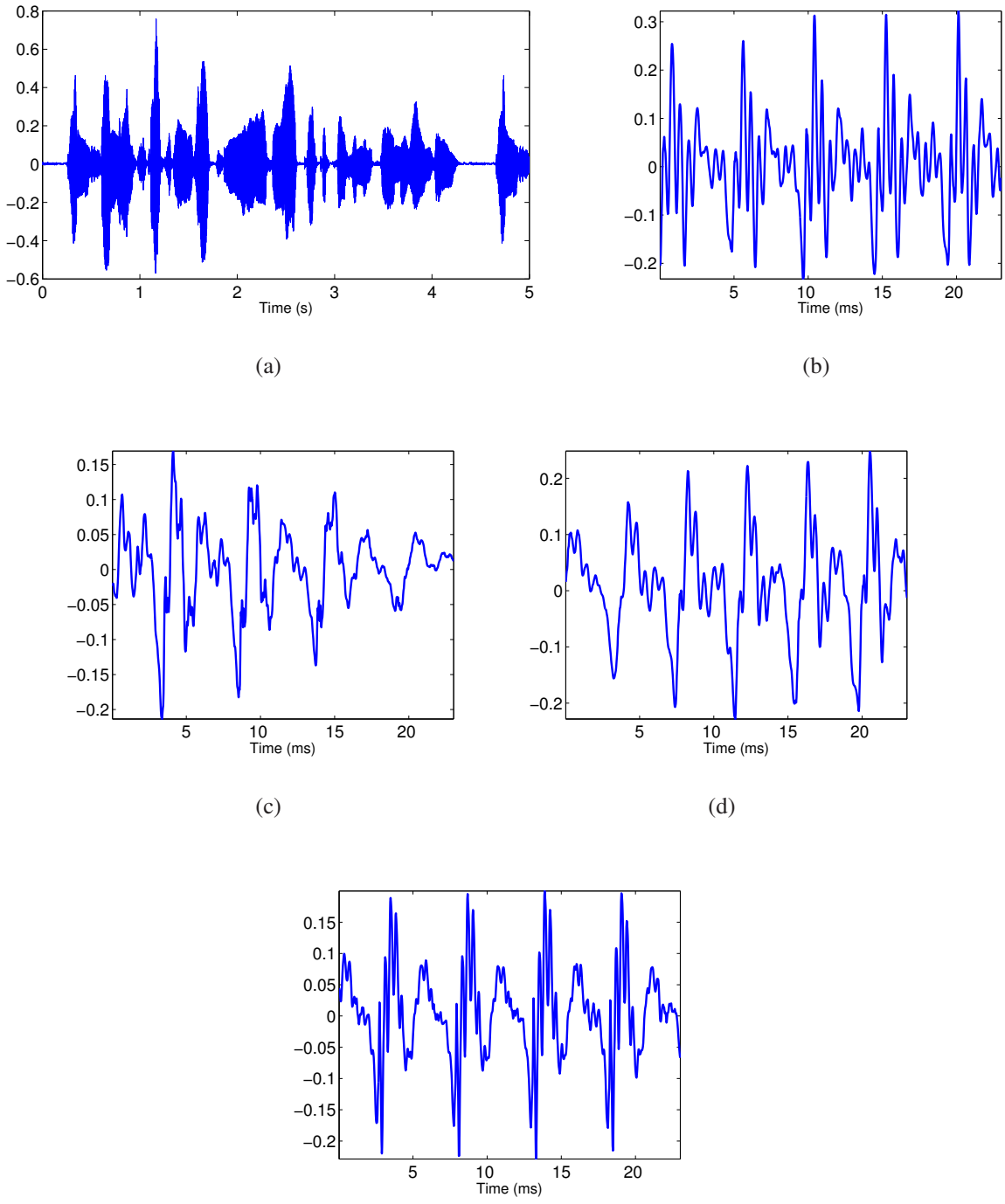
One of the successful applications of this approach belongs to speech processing schemes. Due to the slow changes in the human's vocal system, the speech signal can be considered stationary over short segments (usually every 10 to 30 milliseconds). The example shown in Fig. 1.6 supports the short-term stationary assumption in speech processing. A 5 s speech signal displayed in Fig. 1.6(a) belongs to a female speaker. Four 23 millisecond frames are randomly selected from this speech signal as shown in Figs. 1.6(b) to 1.6(e). It can be seen from the plots that although the long speech signal is non-stationary, the short frames are mostly periodical, and can be assumed stationary.

While short-term analysis have been successfully used in several applications, there are two main shortcomings as follows:

- Short-term analysis might segment the signals into parts that may not be considered stationary. Therefore, this approach fails to accurately track the non-stationary structures of signals. This is a serious problem particularly in the cases of real-world applications where signals contain abrupt changes or discontinuities.

- The short-term analysis approach does not use the long-term information hidden in a signal. In manual data screening and decision making, the human perception requires at least a few seconds to process a given data to better understand the underlying information. For example, an average individual has to listen to at least 0.5 s to 1 s of an audio to understand the content, and even to learn more about the details in the audio, the person requires to listen to a longer episode of that signal. Similarly in automated signal processing, when a longer signal is analyzed, a better understanding of the signal characteristics will be obtained.

On one hand, an automated signal processing algorithm should be performed over long segments of the signal. On the other hand, short-term analysis is effective only on short and stationary durations of signal. This trade-off inspires us to focus our attention on developing methodologies that can be applied to long-term and non-stationary signals without any stationarity assumption about the signal.



(a) (b) (c) (d) (e)

Figure 1.6: (a) A 5 s speech signal with sampling frequency of 44.1 kHz. (b) to (e) 23 millisecond segments randomly selected from the speech signal in (a).

1.4 Structure of Automatic Signal Analysis

The general purpose of any automatic signal analysis technique is to analyze a given dataset and make a certain decision based on the obtained information.

1.4.1 Pattern Recognition

Pattern recognition aims to classify a given signal based on the observations from the data. Depending on the application, these measurements or observations are collected based on either prior knowledge or a set of statistical information extracted from the data. The block diagram in Fig. 1.7 shows the four stages exist in a complete pattern recognition system. The first block includes a sensor that gathers the observations to be classified or described. Sensors measure physical quantities and convert them into signals which can be recorded for further analysis. Some examples of sensors include: thermometer for measuring temperature, microphone for recording audio and speech, and acoustic and pressure sensors for collecting biomedical signals. Depending on the sensor, the translated quantity may be produced in different properties of its output signal. For instance, the amplitude of signals generated using microphone or thermometer transducers represent the measured quantity, while the phase or frequency of the signals created using acoustic-based transducers correspond to the quantity of our interest.

The second block in Fig. 1.7 is the signal preprocessing which contains one or two signal processing stages that provide an optimum representation of the signal. This stage could include simple noise reduction or domain transformation (i.e. switching from time domain to frequency or joint TF plane.). The third block is feature extraction that maps the signal into an appropriate multi-dimensional space. The final block is a classification scheme that performs the actual task of classifying the signals relying on the extracted features.

One of the main challenges in a pattern detection system is to detect distinguishing signatures of a given signal among the other signals. However, depending on the application in hand, the distinguishing pattern, also called the pattern of interest, can be categorized in one of the following scenarios:

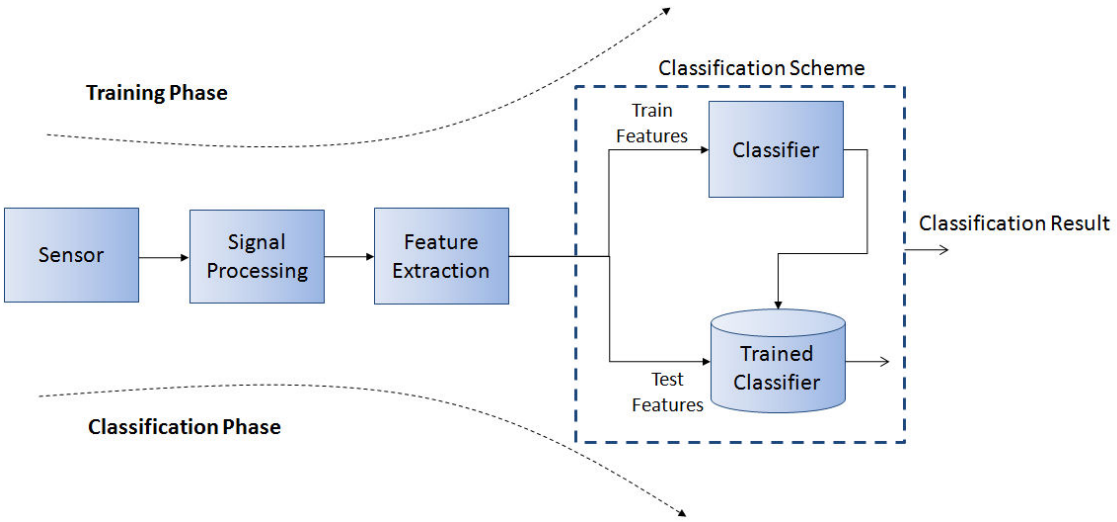


Figure 1.7: Schematic of a complete pattern recognition system.

- **Known structures**: There are situations in which the characteristics of interest are known to us, and our objective is to detect the presence of the known structure. Although this might sound like a straightforward signal detection problem, the application of such scenarios in real-world signals can be a very complicated task. Detection of a face in a surveillance system is an example of a known pattern detection where the face to be searched is the known pattern, and the objects and other faces in that image are outliers. Although in this example the discriminant pattern is known, the presence of outliers and non-stationarities in the pattern makes the detection process very challenging.

- **Embedded structures**: In this scenario, the patterns of interest are intentionally embedded into a signal, and our goal is to detect their presence. The main application of such deliberately embedded signals is in multimedia security, in which a known signature is hidden in the host data such that the added message secures the content. Although the patterns of interest are known to us, the main difference between the known and the embedded structure is that the latter one deals with the embedded patterns which do not belong to the signals' nature and are externally added to the signal, while the former works with the patterns that originally belonged to the data.

- **Unknown structures**: Some applications require to first reveal the distinguishing characteristics of the signal, and then classify the signal based on the existence or the absence of that pattern in the signal. This is the case in most signal classification problems where the discriminating pattern is unknown and complex so it cannot fit in a pre-known structure. Unlike to known or embedded structures, the pattern recognition system in this scenario has to first identify the patterns of interest before it can develop any model or make any decisions about the data.

In this dissertation, our goal is to improve the performance of a pattern classification system as related to all the above scenarios.

Illustration of Pattern Recognition

Fig. 1.8 describes feature extraction and classification schemes in a toy example. Two classes are generated: G_1 and G_2 with 10 signals in each class. The signals are generated using a Gaussian function with mean and variance of μ and σ^2 as defined below:

$$y = \frac{1}{\sigma\sqrt{2\pi}}\text{exp}-\frac{(x-\mu)^2}{2\sigma^2} \tag{1.1}$$

The mean and variance of each class are different; G_1 contains Gaussian signals with mean and variance of (0,1), while mean and variance in G_2 signals are (0,2), respectively. Figs. 1.8(a) and 1.8(b) show one sample signal from each class.

The goal in pattern recognition is to find a discriminating pattern between G_1 and G_2 signals, and then use this discriminating pattern to decide whether the new signal belongs to G_1 or G_2. To make this happen, we calculate the variance of each signal; circle and square points in Fig. 1.8(d) represent variance of all the signals in class 1 and 2, respectively. From the obtained features in Fig. 1.8(d), it can be observed that all the signals belonging to class 1 have variance values of less than 1.2, while all class 2 signals have values of more than 1.4. Based on this observation, we define a classifier which classifies any signal with variance of less than 1.25 in class 1; otherwise the signal is assigned to class 2. To test the pattern recognition system, we generate a test signal using the same method as in G_2 signals. The new signal is shown in Fig. 1.8(c), and its variance

points is displayed as a star in Fig. 1.8(d). The point is above the classifier line, therefore, the test signal is correctly classified as class 2.

The above example demonstrated a successful pattern recognition scheme in which variance is the extracted feature, and the line at 1.25 is the classifier. The following sections further explain feature extraction and classification schemes.

1.4.2 Feature Extraction

Feature extraction involves simplifying the amount of resources required to accurately describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with a sufficient accuracy. As mentioned, features play a very important role in any pattern recognition system. If the extracted features are so well defined, even simple classification methods will be good enough to accurately and efficiently classify the data. Therefore, developing more powerful features and understanding the feature space should be a vital consideration in designing automatic decision making algorithms.

In order to better understand the importance of feature extraction, let us change the features in the Gaussian example shown Fig. 1.8. We replace the variance-based features with new features which are chosen to be the mean of each signal. Fig. 1.9 displays the new feature space with average of each signal as the signal representative. It can be seen in this figure that unlike the variance-based features, the new features are overlapping, and it is impossible to define a linear classifier to separate the classes in the new feature domain. Although both mean and variance are the signals' statistics, the former cannot discriminate the two classes, while the latter can successfully do so.

As demonstrated in the above example, the right features led to a successful pattern recognition system. However, if the extracted features are not appropriately selected, the signals can not be discriminated in the feature domain, and as a result the system will fail to find a discriminating

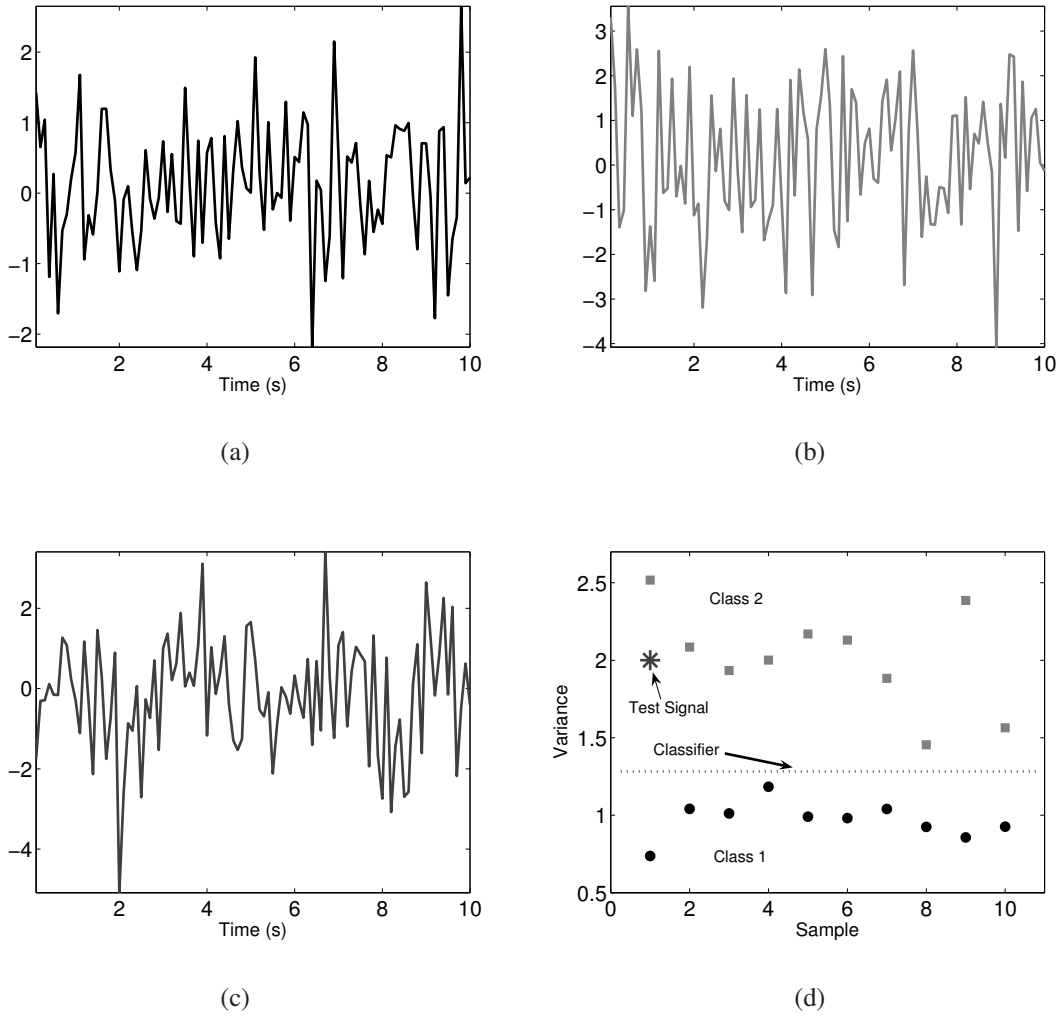


Figure 1.8: (a) A signal in Class 1 (G_1). (b) A sample from Class 2 (G_2). (c) Test signal. (d) Feature plane; each point represents a signal in feature plane.

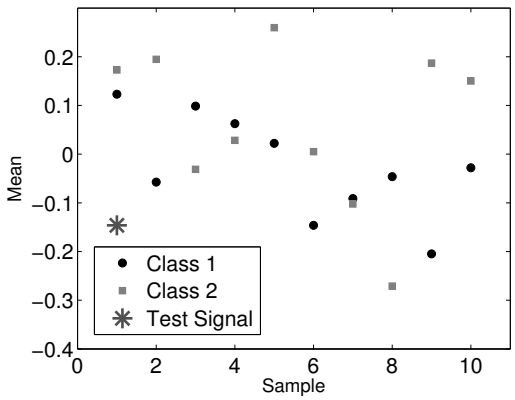


Figure 1.9: Feature plane of the example shown in Fig. 1.8. Instead of variance, average of each signal is used as representative features.

structure between the signals.

In general, the powerful features constrain the following properties:

- **Representative:** Desirable features represent the common characteristics that exist among the signals belonging to the same class. Let us take the feature extraction illustration above as an example. In this example, the variance and mean are among the common properties of the signals in each class.

- **Discriminative:** Powerful features are not only representative of the signal pattern in each class, but also discriminative of each class from the other classes. Referring to the example of Gaussian signals with the same mean but different variances, it was evidenced that only the variance-based features could discriminate the two classes.

- **Localized:** Features are preferred to locate the discriminant pattern in both time and frequency. This property is predicted to not only increase the decision making accuracy, but also locate the pattern of interest. Localized features could have substantial benefits in many pattern classification applications. For example, in knee joint problems, localized features extracted from vibroarthrographic signals emitted during active movement of leg could help us: i) to classify the pathological condition of the articular cartilage; and ii) to automatically

locate the actual problem in order to eliminate the need of painful and costly surgeries for knee disorder diagnosis.

- **Meaningful:** Features can be either meaningful or abstract. The former refers to features which relate to a physical meaning or the generation process of the signal. The latter states those features which do not represent any definable signal characteristic. An example of meaningful features is cepstrum coefficients. Cepstrum coefficients, which have been widely used as speech signal features, approximate the human auditory system's response. Statistical features, such as mean or variance, are examples of abstract features. The majority of features are abstract, which are often extracted by applying complex statistical operations on the signal. Meaningful features are more desirable as they can better relate to the physical changes in the signal due to the non-stationarity. Therefore, these features have the potential to better represent the signal and increase decision making accuracy. Additionally, in biomedical signal analysis, meaningful feature may develop new quantities that can be used for better understanding of the physical behaviour in the human body.

1.4.3 Classification

Classification refers to a prediction rule that assigns the signals into different classes. As shown in Fig. 1.7, classification scheme consists of a training and a classification phase. The former is a computational procedure that trains the prediction rule based on a set of signals which are termed the training set, and the latter classifies any new signal from a testing set. In general, classification techniques can be divided into two groups: supervised learning and unsupervised learning. In supervised learning, the classification scheme is usually based on the availability of a set of signals that have already been classified or described. The resulting learning strategy is characterized as supervised learning. Learning can also be unsupervised, in the sense that the system is not given a prior labeling of patterns. Instead, it establishes the classes based on the statistical or structural regularities of the patterns. The difference between the data labeling in supervised and unsupervised learning is displayed in Fig. 1.10. Fig. 1.10(b) shows the feature space of a data in supervised learning, and Fig. 1.10(a) displays the same feature space in an unsupervised scheme.

In the supervised learning, the labeling of each feature point is known; hence, the feature points are shown with different marks (i.e., asterisk and circle) depending on the corresponding class. Since the labeling information is not known in the unsupervised approach, the points from different classes are not visually differentiated.

Supervised learning might be more advantageous compared to the unsupervised ones in the sense that it uses more prior knowledge about the data. For example, the number of classes and the labeling of each signal are known in the supervised learning, which can result in a more accurate prediction rule for the training set. On the other hand, unsupervised classification is a natural way to proceed towards automatic pattern recognition systems as it provides the automatic clustering of the features in the same way as a human intelligence system is organized. It also gives an insight about the existent structures and patterns in the data. For instance, in the dataset shown in Fig. 1.10(b), two classes are originally reported. Hence, the supervised classifier finds a decision rule based on this given assumption (shown in Fig. 1.10(d)). However, as it can be seen in Fig. 1.10(c), since an unsupervised classifier is not restricted to a certain number of classes (two classes in this example), it is able to accurately find four classes in the dataset. It can be seen that unsupervised learning enables us to obtain more adaptive and meaningful classes corresponding to the natural characteristics of the data.

1.5 Review of Previous Works in TF Analysis

As most practical signals are non-stationary, time domain, $x(t)$, is not enough for representation of the non-stationary signals. Fourier representation, $X(f)$, reveals spectral features of the signal, but it does not preserve any explicit localization in time. On the other hand, joint TF analysis is more suitable for revealing the non-stationary behavior of signals such as trends, discontinuities, and repeated patterns where other signal processing approaches fail or are not as effective [2]. Some TF analysis examples include, but are not limited to, the work of Duze et al. [3], Williams et al. [4] and Stridh et al. [5] in which the authors introduce the advances of TF representation for visualization of the event of interest in electroencephalographic (EEG) and electrocardiogram (ECG) signals. In general the work in the area of TF analysis can be divided into one of the

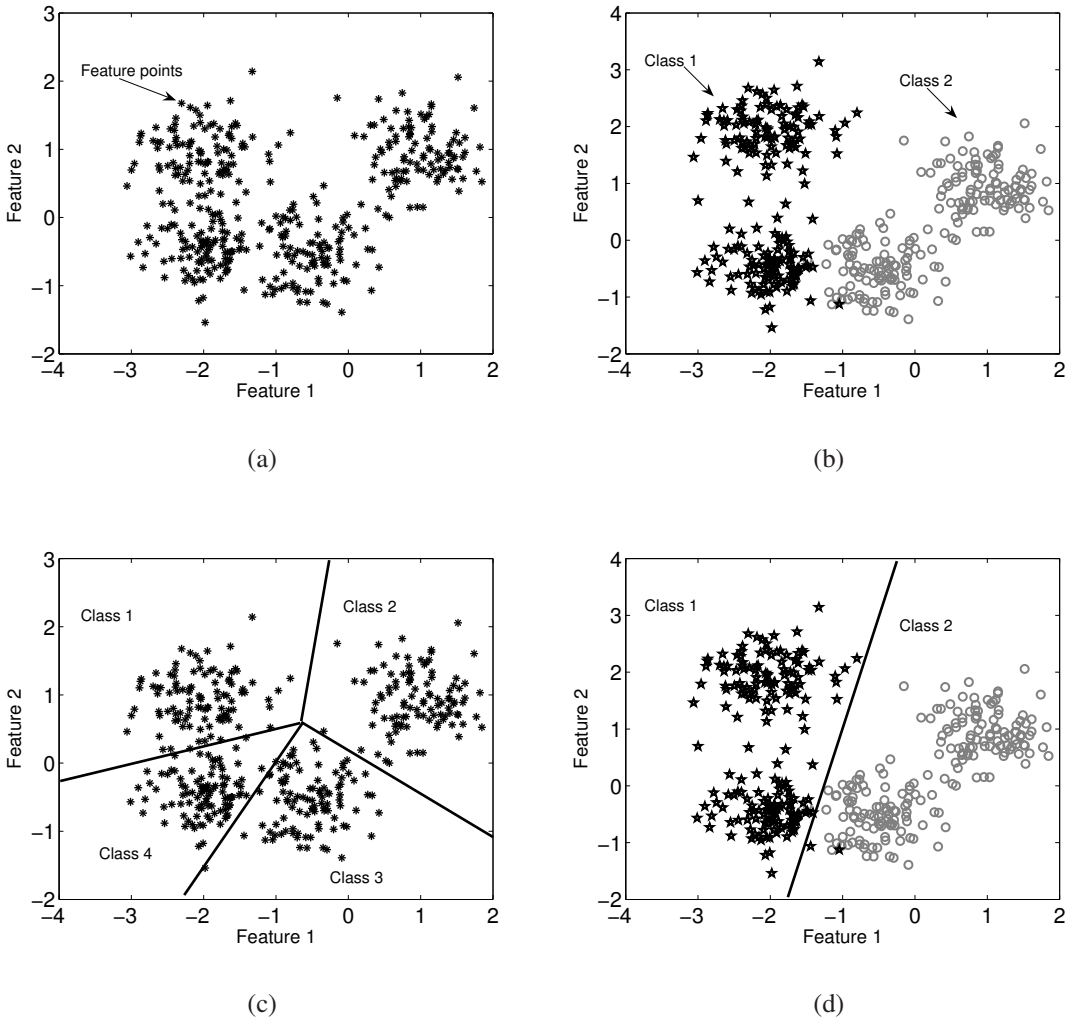


Figure 1.10: (a). A synthetic feature set with no labeling information. (b) The same feature set with known labeling information. Two classes are reported in this dataset; Class 1 data is plotted in star points, and Class 2 data is represented by circle points. (c) An unsupervised learning method divides the data into four classes by dividing the feature plane into four regions. (d) A supervised learning divides the feature plane into two classes. The unsupervised learning obtains more adaptive and meaningful classes corresponding to the natural characteristics of the data.

following categories: i) visualization, or ii) quantification.

1.5.1 TF Analysis for Visualization

Recently, there has been a growing interest in TF analysis for the purpose of representation of the event of interest. For instance, in [6], Williams et al. use TFD to represent the event-related potential (ERP) activities. In another work, Delorme et al. [7] use TFD of multi-channel EEG signals for visualization of the temporal dynamics of the brain activities and interactions. Morup et al. [8] adopt TFD to visualize the inter trial phase coherence (ITPC) of multichannel EEG. Rutkowski et al. in [9] perform TF analysis of multichannel EEG signals to find and enhance very small oscillations related to presented visual stimulation.

While the above literatures are very beneficial in the area of visualization of the event of interest, there are two major differences between the above studies and this dissertation as follows:

- The above analysis schemes mainly focus on visualization enhancement and do not include any classification and decision making process. The main objective of the above studies is to represent the event of interest in the recordings, and therefore, they restrict the TF analysis to visualization of the event of interest. However, in the present dissertation, we focus on TF quantification for pattern recognition.

- In the above studies, the authors mainly consider the TF analysis of a multi-channel data, while the present dissertation focuses attention on one channel TF analysis where can be also extended to multi-channel signals. For example, in [6, 7], the TFDs of all the channel ($\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_K$) are transformed into one TF matrix (TFM) ($\hat{\mathbf{V}}$), and then the constructed TFM is used for further processing. However, in the present study, we consider the TF quantification of one waveform, where the entries in the TFD denote energy values with rows denoting frequency and columns denoting time.

1.5.2 TF Analysis for Quantification

While the benefit of TF representation for visualization purposes is certain, because of the high dimensionality of TF plane, the direct application of the TF distribution in a pattern classification

system is not efficient. Lately, there have been some attempts to reduce the dimensionality of the feature space by removing the redundancy and keeping only the representative parts of the TFD [10, 11, 12, 13, 14, 15, 16, 17], [18, 19].

Different TF quantification notions have been introduced in the literature. In some approaches [11, 13], TFD is constructed as a 2-D probability density function (pdf) of the signal's joint time and frequency behavior, and some representative statistics are extracted from the constructed TF pdf. Although this approach decreases the dimensionality of the TFD to some extent, the signal is still represented with a large number of TF features. In some of the other approaches, [15, 16, 17], the TFD is interpreted as a matrix denoted with $\mathbf{V}$. Then a matrix decomposition (MD) technique is applied to the TF matrix (TFM), $\mathbf{V}$, to decompose the TFM into its decomposed matrices, $\mathbf{W}$ and $\mathbf{H}$, in a way that $\mathbf{V} \approx \mathbf{WH}$. The decomposition is performed in a way that $\mathbf{W}$ contains the spectral structures, and $\mathbf{H}$ contains the corresponding temporal location of each spectral structure in the TFM. In this approach, the decomposed matrices are used as TF feature vectors. The problem with this technique is that the dimension of extracted feature vectors are still very large. This is because the length of each feature vector is proportional to the signal's sampling frequency, and as a result they are not very appealing for classification applications. There are some TF quantification methodologies [20, 21] which are a combination of the first two approaches. First they use a MD to reduce the TFM into its spectral and temporal vectors, and then they decrease the decomposed vectors' dimensionality by considering each vector as a pdf, and extracting the statistical features.

The above TF quantification notion is effective if the obtained TF features represent the discriminative TF structures of the signals. Identifying the discriminant TF structure has attracted attention in literature. Local discriminant base (LDB) analysis [22] is a wavelet packet based approach to identify the discriminative bases in the TF plane. While LDB analysis and its variants are an active area of research [23, 24, 25, 26], the optimal choice of LDBs highly depends on the nature of the dataset and the dissimilarity measures used to distinguish between classes. In another work, Umapathy et al [27] proposed a time-width versus frequency band (TWFB) energy mapping to visualize the discriminative structures within parametric TF decompositions.

In most of real-world applications, the nature of signals from different classes are very similar,

and there is only a slight change in the pattern of one class compared to the other classes. Therefore, the TF bases obtained from TFM decomposition of different classes contain the common bases, which are related to the similar TF structures, as well as to the discriminative TF bases. Since we extract the features from the decomposed TF bases, if we distinguish the discriminative TF bases then the derived features are more suitable for further classification. Having said this, our main goal in this dissertation focuses on TF quantification techniques that achieve representative and discriminant features that potentially improve the pattern recognition performance.

So far in this chapter, we discussed the importance and challenges of non-stationary signal analysis as related to automatic decision making. Furthermore, we introduced TF analysis as a suitable signal representation that provided a comprehensive information in real-world signals and their non-stationary behavior. Also, we highlighted the importance of long-term signal analyses that are adaptive to the non-stationary structures in real-world data. Moreover, we explained that the accuracy of automatic decision making systems depends on developing representative and discriminative features and understanding the feature space.

The challenges involved in analyzing and extracting features from the non-stationary signals, and the absence of a unified approach to automatically extract discriminant features inspires the research of this dissertation. Therefore, the main focus of the present dissertation is to develop adaptive TF analyses that obtain powerful TF features from discriminative areas of signals with different background.

1.6 Contributions of The Dissertation

This work presents a generalized TF analysis methodology that exploits the benefits of TFD in pattern classification systems as related to discriminant feature detection and feature classification. We investigate three different implications of feature analysis as follows: i) detection of known TF features; ii) detection of embedded TF features; and iii) identification and classification of

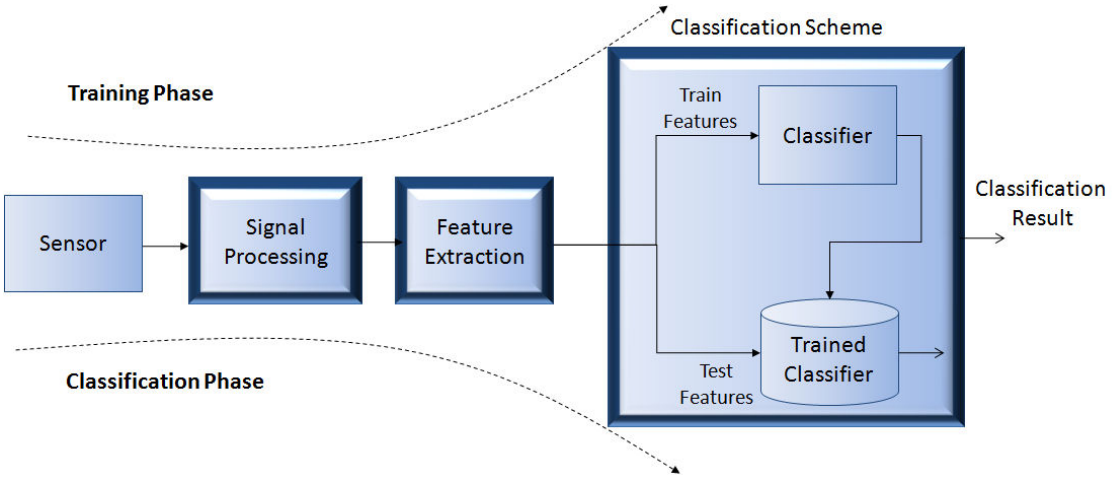


Figure 1.11: General block diagram of the contributions.

unknown TF features. The first two implications focus on the detection of known patterns under the presence of noise and data non-stationarities. The third implication covers wider applications as it aims to classify the signals with unknown discriminant patterns. In all three implications, our main objective is to develop techniques that successfully quantify the patterns of interest. The block diagram in Fig. 1.11 shows the overview of the proposed framework. In this block diagram, three contributed areas are highlighted as explained below.

Signal Processing:

Our main contribution in signal processing stage focuses attention on developing an adaptive and discriminative TF analysis. To fulfill this objective, in the first point, we intend to build a time-frequency decomposition technique that increases the effectiveness of segmentation in real-world signals. In the second point, based on the above technique, our goal is to develop a unique and novel discriminant TF (DTF) analysis method to perform automated and discriminative feature selection of any non-stationary signals. The desirable DTF analysis automatically identifies the differences between different classes as the distinguishing structure, and uses the identified structure to accurately classify and locate the discriminant structure in the signal.

Feature Extraction

In the next step, we intent to extract meaningful and unique features from the data. To make this happen, once the main TF components of a signal are identified based on the above signal processing technique, we introduce novel features in a way that they represent the TF structure of the signal. Additionally, the developed method makes sure that the obtained features are robust to noise and outliers and are effective for classification and detection of the discriminant patterns in signals.

Classification

We seek a new discriminant clustering approach that improves the classification accuracy in decision-making systems. As mentioned in this chapter, in many applications, the nature of signals from different classes are very similar. However, current approaches assume that the structures of signals from different classes are completely different, and as a result, the obtained features from those classes might be overlapped in feature space. The proposed feature clustering technique is aimed to solve this problem by developing a new machine learning approach. This approach considers the fusion of supervised and unsupervised classification methods to cluster the features that represent the discriminative pattern in a given data. The discriminative clusters are then used to compute the presence of the discriminative pattern in any given signal. Such a framework can significantly improve the classification accuracy rate of the signals.

1.7 Organization of The Dissertation

The dissertation is organized in 9 chapters. The flowchart in Fig. 1.12 displays the evolution of this dissertation. The objectives at each stage are shown at the bottom of this chart. We begin with identifying the right TFD technique for the proposed work that best suits the non-stationarity in complex real-world data. In Chapter 2, the appropriate TF transformation is selected. Three TF quantification approaches are investigated: i) detection of known TF features; ii) detection of embedded TF features; and iii) identification and classification of unknown discriminant TF features. The first two problems are studied in Chapters 3 and 4 respectively, and the third problem is

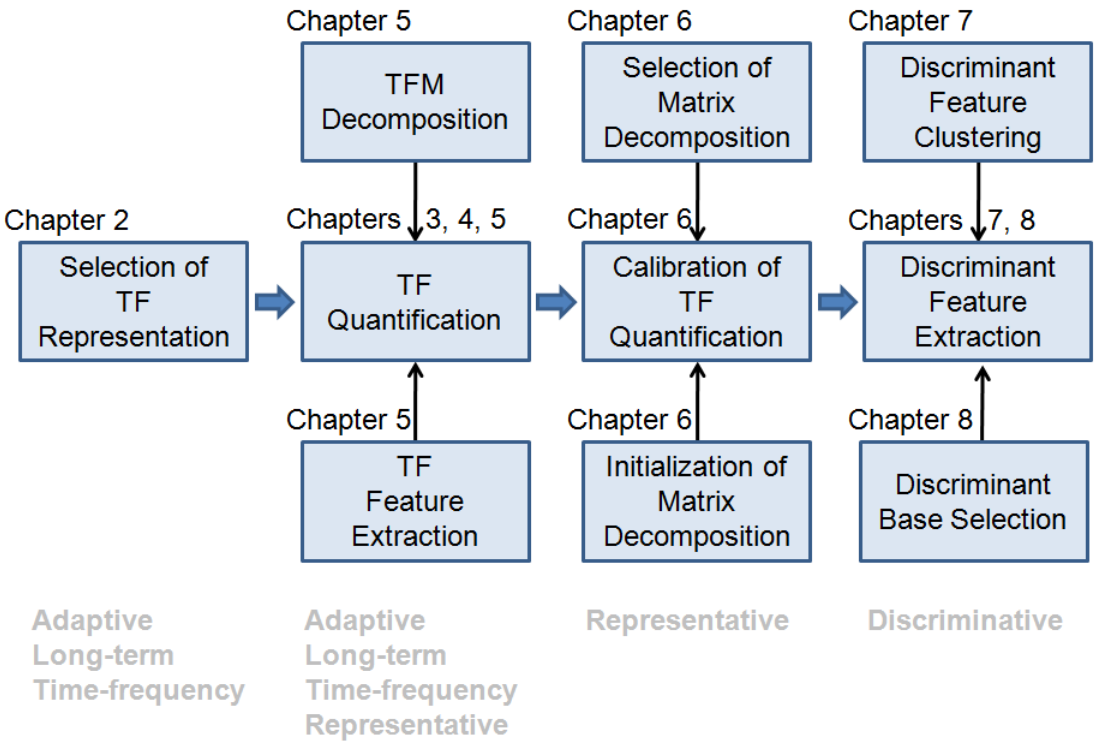


Figure 1.12: Flowchart of the proposed contributions.

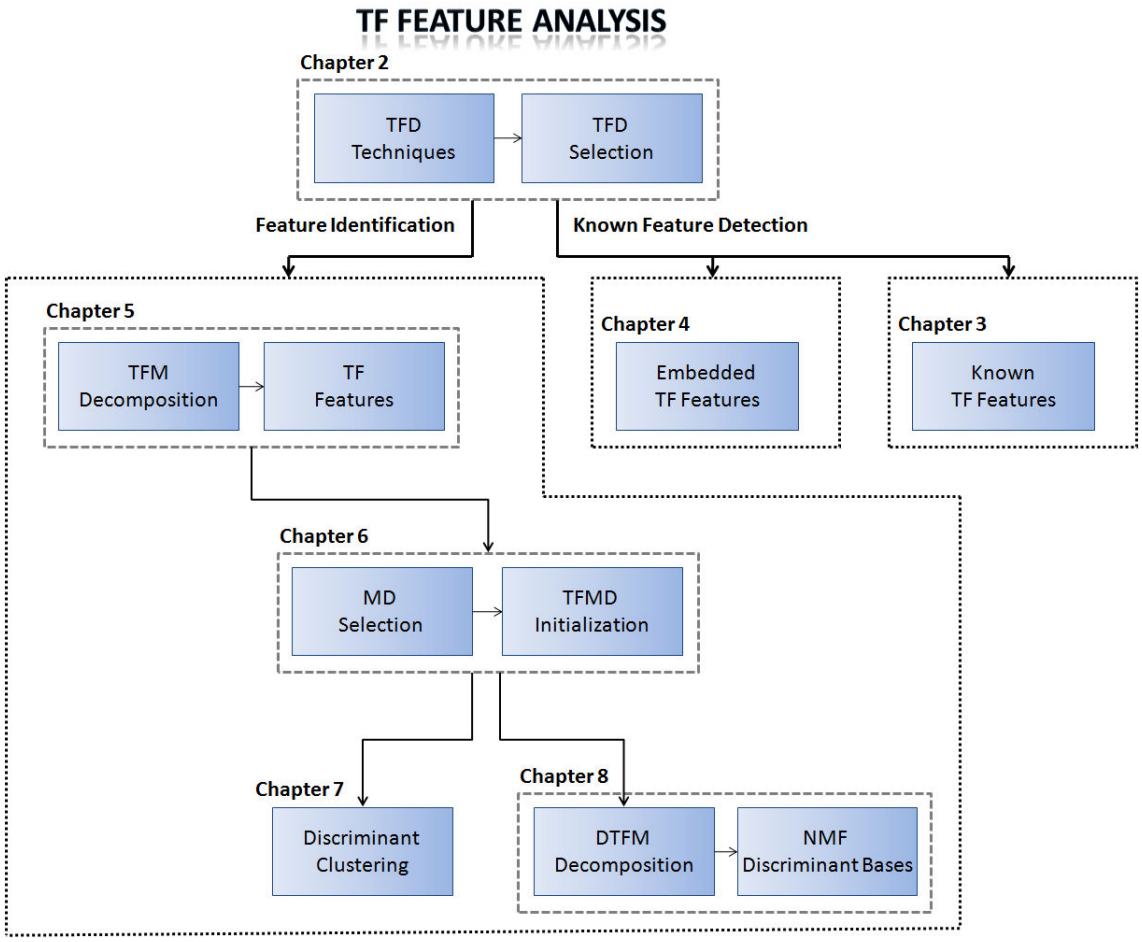


Figure 1.13: Block diagram of the dissertation.

investigated in Chapters 5 to 8. In Chapter 5, a novel TF quantification methodology is developed. The new framework consists of two stages: TF decomposition and TF feature extraction. The developed TF quantification is calibrated in Chapter 6 where we select the right tools for TF decomposition. This stage makes sure that the derived features provide significant representation in order to achieve an enhanced pattern recognition system. In Chapters 7 and 8, we further develop the framework so that the extracted features are more discriminative and as a result more suitable for classification of real-world signals. Wherever possible the chapter includes an experimental investigation in addition to the analytical and algorithmic frameworks.

The block diagram in Fig. 1.13 shows the organization of the dissertation.

Chapter 2: Time-frequency Representation

This chapter covers the time-frequency theory and discusses the existing TF representation methodologies in terms of their achievable TF resolution and suitability for the proposed work. Synthetic examples and TF distributions are used in explaining the TF properties. In this chapter, the TF representation tool that satisfies the requirement of efficient non-stationary signal analysis is chosen.

Chapter 3: Known TF Feature Detection

The problem of TF feature detection as related to known signal patterns is explained in this chapter. In some applications, the structures of interest are known, and our objective is to quantify the known pattern into representing features that are robust to signal non-stationarities and outliers. Once the desirable features are obtained, we use them to detect the pattern of interest in any given data. In this chapter, adaptive time-frequency quantification is introduced as a successful tool to quantify and effectively track such signatures in a given signal. The proposed technique is utilized to enhance the procedure that identify patients with heart disease who may experience sudden death from ventricular arrhythmias.

Chapter 4: Embedded TF Feature Detection

The objective of this chapter is to quantify the known TF patterns that are deliberately embedded in a signal. In the previous chapter, we studied the known structures that originally belonged to the data. However, there are scenarios where the pattern of interest belongs to an external signature which is intentionally inserted into the data. Our goal in such applications is to develop signal quantification techniques that are invariant to the non-stationarities in the pattern of interest. The proposed technique intends to extract TF features that effectively quantify the time and frequency varying structure of such signatures in the TF plane. Multimedia security was presented as a real-world signal example demonstrating that the proposed TF quantification technique improves the pattern detection even in the presence of noise and signal manipulations.

Chapter 5: Time-frequency Quantification

This chapter introduces a novel TF decomposition technique that adaptively decomposes a non-stationary signal. The proposed technique which is called TF matrix (TFM) decomposition is a window-less approach that is applied to the entire data without any need to blind segment the data into short durations. Additionally, this chapter proposes a TFM quantification that preserves the time and frequency localization of a given signal and provides a significant low-dimensional and yet powerful quantification tool for real-world signals. The performance of the proposed TF quantification methodology is demonstrated through some synthetic and real signals.

Chapter 6: Matrix Decomposition Analysis

This chapter covers and discusses the existing matrix decomposition (MD) techniques in terms of their achievable decomposition accuracy and suitability for the TF quantification. Other contributions of this chapter includes integration of TF transform with MD optimization to achieve a faster and improved convergence of the algorithm. We apply the developed novel TF quantification methodology to audio scene classification with a diverse database, and detection of the risk of sudden cardiac death (SCD) in patients with heart problems.

Chapter 7: Discriminant Feature Selection

To enhance the discriminatory power of the extracted TF features, this chapter presents a novel machine learning approach to select the discriminant TF features obtained from the decomposed TF components. This approach flexibly selects the feature points according to their importance in representing the patterns of interest. Synthetic and real-world examples demonstrate the applicability of the developed TF feature selection approach. Audio scene analysis and pathological speech classification problems are explained to verify the efficiency of the proposed work.

Chapter 8: Discriminant Bases Selection in TF Matrix Analysis

In this chapter, we propose a novel discriminative TF quantification method that adaptively identified the long-term and discriminant TF structures between two signals to improve the detection

accuracy of discriminant structures. The formulation and theoretical properties of the novel TFM decomposition pertaining to pattern recognition are presented using synthetic signal examples. Synthetic signal examples and real world examples are presented in verifying and demonstrating the utility of discriminant base selection method in identifying the discriminative structure.

Chapter 9: Conclusions and The Future Work

A summary of the complete work is presented with analysis of the achieved results at various stages. The novelty and the multifold benefits of the proposed work is highlighted. A discussion on the potential of the proposed methodology in formating as a versatile non-stationary signal analysis tool and the future directions on enhancing the same are presented.

Chapter 2

TIME-FREQUENCY REPRESENTATION

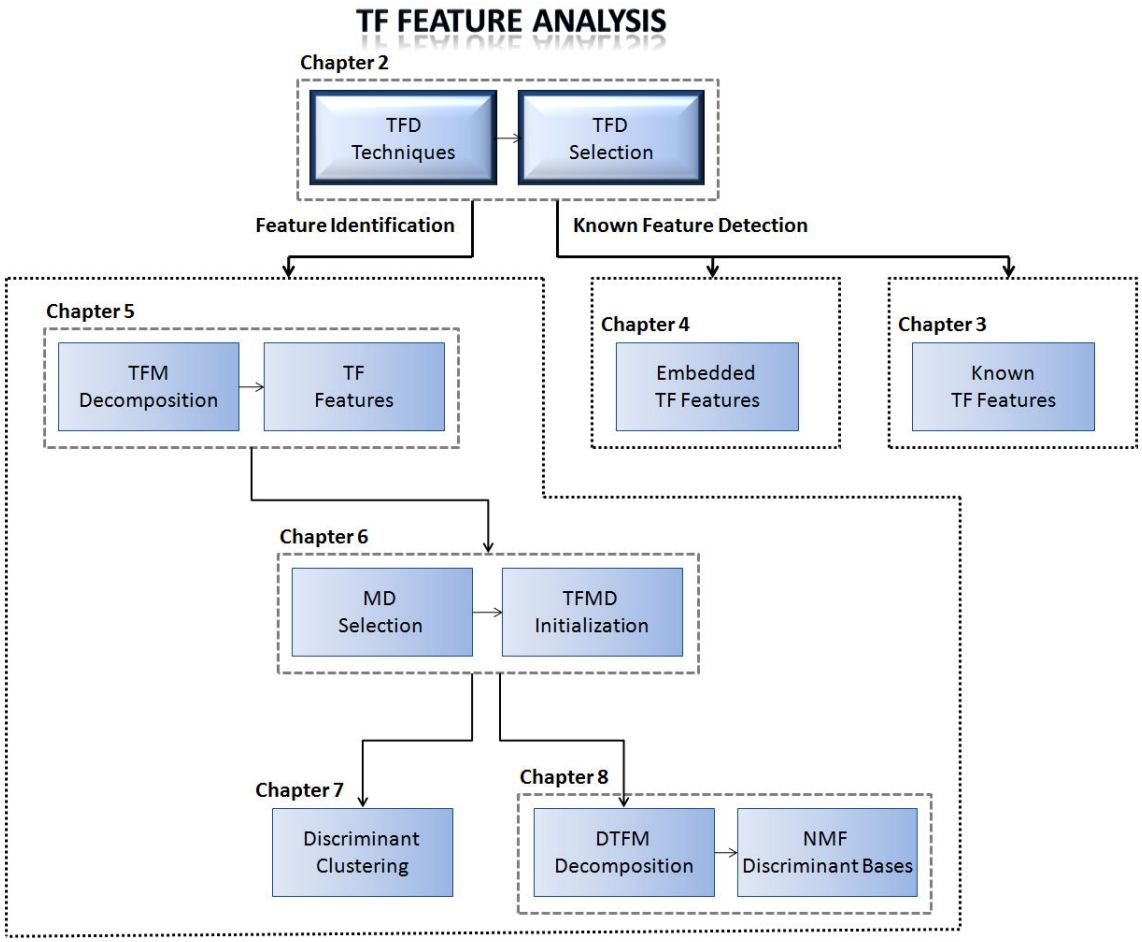


Figure 2.1: Chapter 2 - Selection of TF representation.

PRACTICAL signals are non-stationary, and they therefore can not be efficiently represented in the time domain, $x(t)$. Fourier representation, $X(f)$, reveals spectral constitutive features of the signal, but it does not preserve any explicit localization in time. It is well-known that Fourier representation faces limitations when we are looking for non-stationarity features of the signal. Hence neither time-domain nor frequency domain analysis are sufficient enough to analyze signals with time-varying frequency content. To overcome this difficulty and to analyze the non-stationary signals effectively, techniques that provide joint time and frequency information are needed. Joint TF distribution (TFD) indicates a two dimensional energy representations of a signal in terms of time and frequency domains.

The work in the area of TFD methods is extensive [28, 29, 30]. Depending upon the application in hand and the feature extraction strategies, any of the TF approaches could be used. Therefore, in this chapter, we select the TFD that is most appropriate for characterization of non-stationarities as related to pattern recognition. Fig. 2.2 demonstrates the structure of this chapter. First, we explain the well-known TFD methods. Next, we describe the desirable properties in terms of their achievable TF resolution and suitability for pattern recognition. Finally, based on these properties, we select an appropriate TF transformation.

2.1 Time-frequency Distributions

Fig. 2.3 displays the most well-known TF transformation techniques. Any of these techniques transform a temporal signal, $x(t)$, into a TF distribution denoted with $\mathbf{V}(t, f)$, where t and f locate each sample in the TF plane and $\mathbf{V}(f, t)$ is the TF value at the corresponding location. Such transformation is displayed in Fig. 2.3.

Depending on the transformation technique, TF distributions with different properties are achieved. For example, some methods construct distributions with non-negative entries while some might result in negative values also. Temporal and Spectral marginals of a TFD are calculated along each time and frequency coordinates as shown below:

$$TM(t) = \int_{-\infty}^{+\infty} \mathbf{V}(t, f)df, \tag{2.1}$$

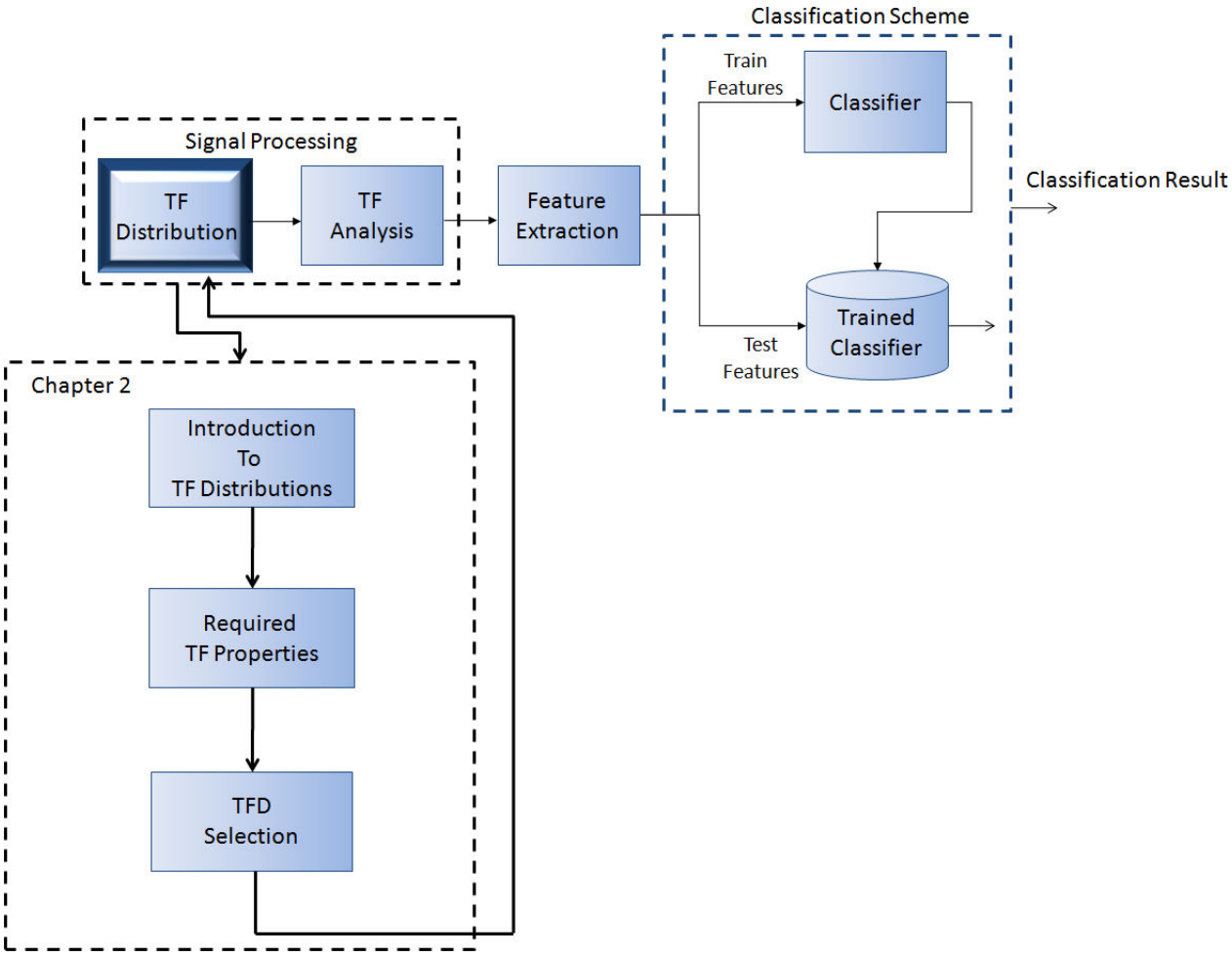


Figure 2.2: Chapter 2 - Selection of TF Distribution.

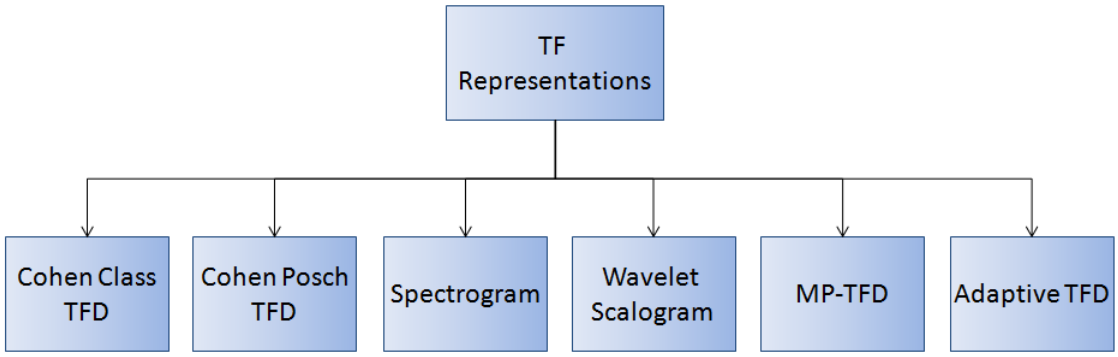


Figure 2.3: A diagram of well-known TF distributions.

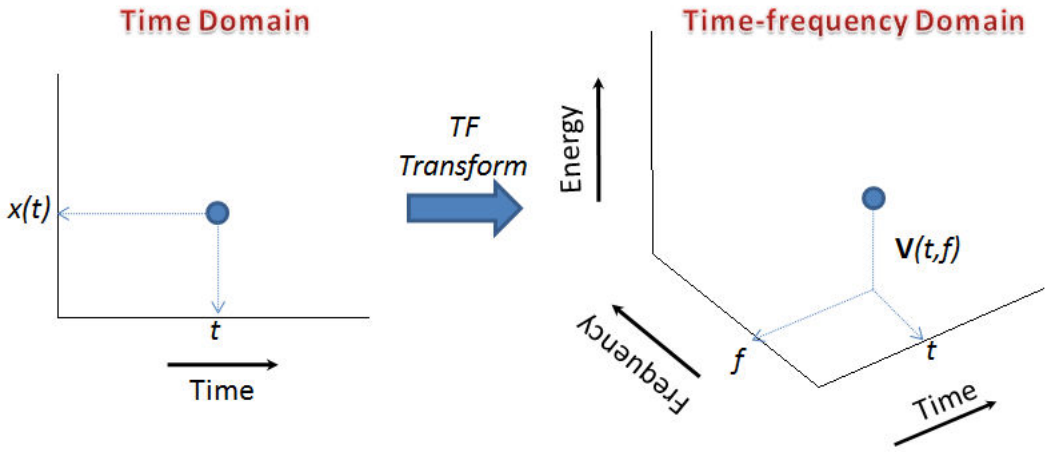


Figure 2.4: A diagram of TF transformation.

$$SM(f) = \int_{-\infty}^{+\infty} \mathbf{V}(t, f)dt, \tag{2.2}$$

It will be discussed later in this chapter that how the non-negativity of a distribution, and its temporal and spectral marginals lead us to achieve a better TF quantification.

2.1.1 Cohens Class Bilinear TFDs

Quadratic methods of TFD will adapt the analyzed signal as the analysis window, i.e quadratic TFD transforms the time varying autocorrelation of the signal to obtain a representation of the signal energy distributed over time and frequency:

$$\mathbf{V}_{WV}(t, f) = \int_{-\infty}^{+\infty} x(t + \frac{1}{2}\tau)x^*(t - \frac{1}{2}\tau)e^{-j2\pi f\tau}d\tau. \tag{2.3}$$

where, $\mathbf{V}_{WV}$ is Wigner-Ville distribution (WVD) of the signal. An example of a WVD is shown in Fig. 2.5.

2.1.2 Cohen-Posch TFD

Cohen-Posch TFD, or positive TFD (PTFD), produces a TFD with non-negative entries. Cohen and Posch [31] demonstrate the existence of an infinite set of positive TFDs, and developed formu-

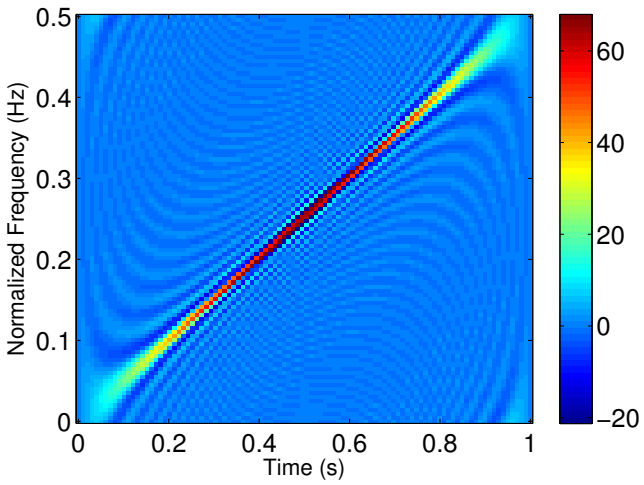


Figure 2.5: WVD of a chirp signal with sampling frequency of 100 Hz and frequency increasing linearly from 0 to 50 Hz.

lations to compute the positive TFDs based on signal dependent kernels as given in the following equation:

$$\mathbf{V}_{PTFD}(t,f) = |x(t)|^2 \, |X(f)|^2 \, \{1 + c\rho \, (s(t), S(f))\} \tag{2.4}$$

where

$$s(t) = \int_{-\infty}^{t} \left| x(\tau)^2 d\tau \right| ; S(f) = \int_{-\infty}^{f} |X(\xi)|^2 \, d\xi, \tag{2.5}$$

and

$$\rho(s(t), S(f)) = h(s(t), S(f)) - h_1(s(t)) - h_2(S(f)) + 1, \tag{2.6}$$

In the above equation, $h(s, S)$ is a positive kernel function of the variables s and S, $0 \leq s, S \leq 1$ and normalized to one. $h_1(s)$ and $h_2(S)$ are the marginals of $h(s, S)$ (defined in Eqn. 2.2), and c is a numerical constant in the range of

$$\frac{1}{\max \left(\rho(s(t), S(f)) \right)} \leq c \leq \frac{1}{\min \left(\rho(s(t), S(f)) \right)} \tag{2.7}$$

2.1.3 Spectrogram

Linear TF analysis decompose the signal over a set of basis functions. The simplest linear TF representation is short-time Fourier transform (STFT) of signal, which assumes that the signal is stationary in short durations and multiplies the signal by a window, and takes the Fourier transform on the windowed segments. The basis functions used in Fourier transform are orthonormal cosine functions with varying frequencies. Fig. 2.6(a) displays few of such basis functions, and the below equation explains how STFT is calculated:

$$\mathbf{V}_{STFT}(t,f) = \int_{-\infty}^{+\infty} x(\tau)h(\tau - t)e^{-j2\pi f\tau}d\tau, \tag{2.8}$$

where $x(t)$ is the the signal, and $h(t)$ is the sliding window function. Spectrogram which is nothing but the squared modulus of the STFT is generally used to display the TF energy distribution over the TF plane. This joint representation of time and frequency is able to represent the frequency content for each time segment. Fig. 2.6(b) displays a spectrogram example.

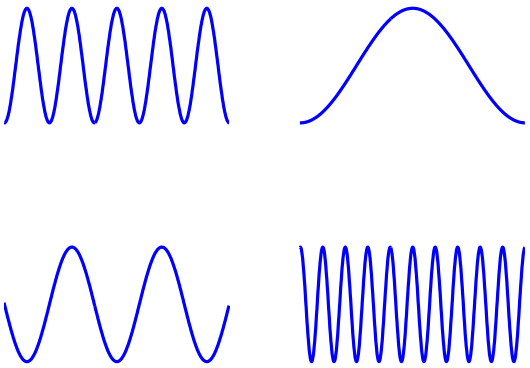
2.1.4 Wavelet Scalogram

Wavelet scalogram is based on wavelet decomposition where orthonormal basis functions with different sizes are used to decompose a signal as given by the following equation:

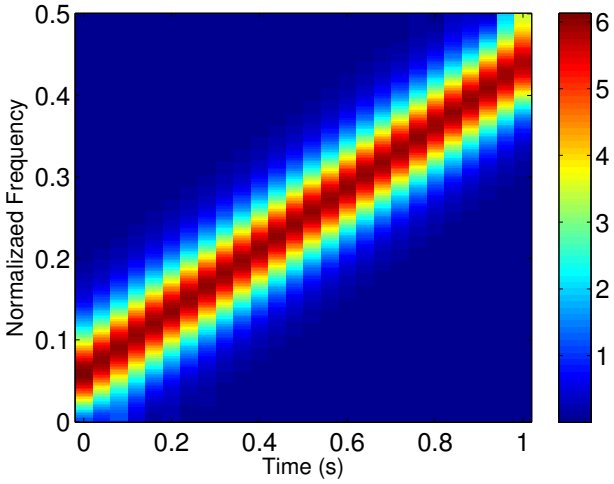
$$\mathbf{V}_{CWT}(t,s) = \frac{1}{\sqrt{s}}\int x(\tau)g(\frac{\tau - t}{s})d\tau, \tag{2.9}$$

where $g(\frac{t}{s})$ is the mother wavelet, and s being the scaling parameter, corresponds to the size of each basis function. In wavelets, the basis function used are small waves called mother wavelets, which satisfy few mathematical conditions. By stretching and compressing mother wavelet, different scaled versions of the mother wavelet are created. These different scaled versions of the mother wavelet are slided across the signals, and models the localized signal structures with the wavelet of a particular scaling. Fig. 2.7 displays Gaussian wavelet functions at different scales.

Wavelet scalogram displays the TF structure obtained from the wavelet transform. In scalogram, each wavelet signal is plotted as a filled rectangle whose its location and size are related to the time interval and the scale range for this wavelet signal. The scaling parameter which stretches



(a)



(b)

Figure 2.6: (a) FFT basis functions at different frequencies. (b) Spectrogram of a chirp signal with sampling frequency of 100 Hz and frequency increasing linearly from 0 to 50 Hz.

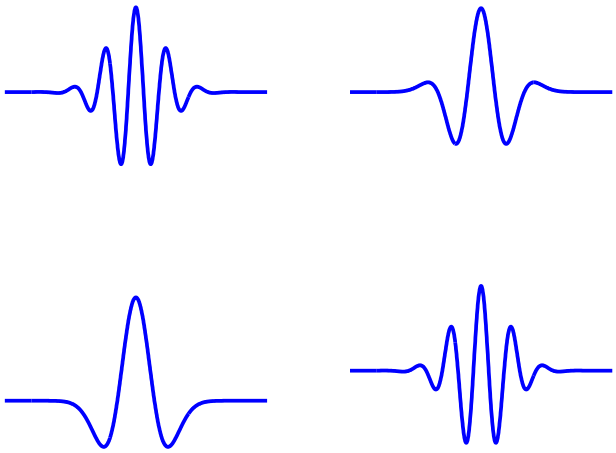


Figure 2.7: Gaussian wavelet basis functions at different scales.

and compresses the wavelets contribute to the change in the center frequency of the wavelets. Small scale factors corresponds to higher frequencies and larger scale factor corresponds to the lower frequencies. In other words wavelets uses short time scales to capture the high frequency structures and a long time scale to capture the low frequency structures in a signal.

2.1.5 Matching Pursuit TFD

Matching pursuit (MP) TF distribution is constructed based on MP decomposition as proposed in [32]. MP uses non-orthogonal basis functions which includes an over-complete and redundant combinations of bases for all possible translations, modulations and scalings as shown in the following equation:

$$G_{\gamma_i}(t) = \frac{1}{\sqrt{s_i}} g\left(\frac{t-p_i}{s_i}\right) \exp\left[j\left(2(\pi f_i t + \phi_i)\right)\right]. \tag{2.10}$$

where $g(t)$ is the primary Gaussian function, and $G_{\gamma_i}(t)$ is a basis function generated from the primary waveform. The scale factor s_i controls the width of the basis function, and the parameter p_i controls the temporal placement. The parameters f_i and ϕ_i are the frequency and phase of the basis function, respectively. The index γ_i represents a particular combination of the TF decomposition

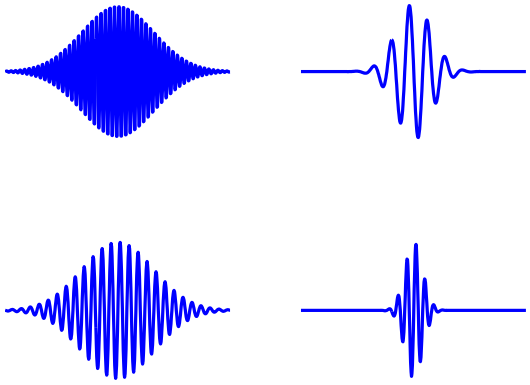
parameters (s_i, p_i, f_i, ϕ_i). The collection of all the possible TF basis functions is called redundant TF dictionary and each member in this collection is denoted as a TF atom. The term redundant indicates that the TF dictionary consists of basis functions much larger than the minimum required orthonormal basis functions to completely decompose a given signal. The dictionary of TF functions is selected based on the application in hand. Since in real world signals, the signal $x(t)$ is real and discrete, we use a dictionary of real and discrete TF functions. The TF dictionary used in this dissertation is the Gabor dictionary which consists of Gaussian atoms of $g(t) = 2^{\frac{1}{4}}\text{exp}^{-\pi t^2}$, which has shown to offer the best TF localization properties [33]. Fig. 2.8(a) depicts Gabor atoms with different scales and frequencies. As it can be seen in the figure, the Gabor atoms are flexible in both frequency and scale.

MP decomposes a signal, $x(t)$, into a linear combination of TF functions $G_{\gamma_i}(t)$ selected from a redundant Gabor dictionary of TF basis functions as given in the following equation:

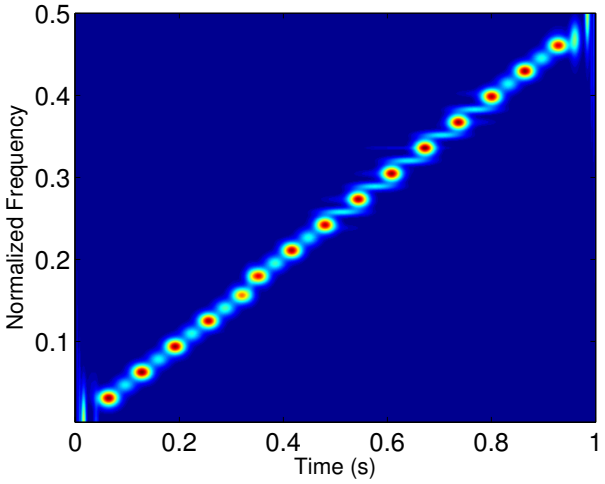
$$x(t) = \sum_{i=1}^{I} a_{\gamma_i} G_{\gamma_i}(t) + R_x^I \tag{2.11}$$

where $x(t)$ being the signal, $a_{\gamma_i} = \left|\left\langle R_x^I, G_{\gamma_i} \right\rangle\right|$ is the expansion coefficient on $G_{\gamma_i}(t)$, and R_x^I is the residue signal after I iteration.

In Eqn. 2.11, the signal $x(t)$ is projected over a redundant dictionary of TF functions with all possible combinations of scaling, translations and modulations. At each iteration, the best correlated TF function was selected from the Gabor dictionary. The remaining signal called the residue was further decomposed in the same way at each iteration subdividing them into TF functions. After I iterations, signal $x(t)$ could be expressed as in Eqn. 2.11, where the first part of Eqn. 2.11 is the decomposed TF functions until I iterations, and the second part is the residue which will be decomposed in the subsequent iterations. This process is repeated till all the energy of the signal is decomposed. At each iteration some portion of the signal energy was modeled with an optimal TF resolution in the TF plane. However, after some iterations, it can be observed that all the coherent structure of the signal is captured in the decomposed components, and the incoherent structure remains as the residue (R_x^I) in Eqn. 2.11. This residue may be assumed to be due to random noise since it does not show any TF localization. Therefore, after high enough iterations,



(a)



(b)

Figure 2.8: (a) Gabor basis functions at different scales and frequencies. (b) MP-TFD of a chirp signal with sampling frequency of 1000 Hz and frequency increasing linearly from 0 to 500 Hz.

the decomposition residue in Eqn. 2.11 can be ignored. The number of required iteration depends on the nature and length of the signal; for example, for 3 s duration of audio signals we found that 1000 iterations are high enough to capture the coherent structure of the signal.

Now that MP selected the collection TF atoms that accurately model the signal $x(t)$, MP-TFD of the given signal, $\mathbf{V}(t,f)$, is constructed by summing the TFD of each decomposed TF atom as shown below:

$$\mathbf{V}(t,f) = \sum_{i=1}^{I} |a_{\gamma_i}|^2 \, \mathbf{WVG}_{\gamma_i}(t,f) \tag{2.12}$$

where $\mathbf{WVG}_{\gamma_i}(t,f)$ is the WVD of the Gabor atom $G_{\gamma_i}(t)$. A MP-TFD example is illustrated in Fig. 2.8(b).

2.1.6 Adaptive TFD

Adaptive TFD method [34] includes an iterative cross-entropy minimization that optimizes the MP-TFD to construct a positive, high resolution and cross term free TFD that satisfies the marginal criteria. This TFD is called Adaptive TFD as it is constructed according to the properties of the signal being analyzed [34].

Cross-entropy minimization is a general method of inference about an unknown probability density when there exists a prior estimate of the density and new information in the form of constraints on expected values is available. In the case of adaptive TFD, MP-TFD, $\mathbf{V}(t,f)$, exists as the initial estimate of the desirable TFD ($\mathbf{V}^{(0)}(t,f) = \mathbf{V}(t,f)$), and its temporal and spectral marginals ($TM(t)$ and $SM(t)$ as derived in Eqn. 2.2) are required to satisfy the following equations:

$$TM(t) = |x(t)|^2 , \tag{2.13}$$

$$SM(f) = |X(f)|^2 , \tag{2.14}$$

The Adaptive TFD is iteratively estimated from the MP-TFD as given in the following steps [34]:

1. The time marginal is satisfied by multiplying and then dividing the TFD by the desired and the current time marginals:

$$\mathbf{V}^{(1)}(t,f) = \mathbf{V}^{(0)}(t,f)\frac{TM(t)}{p^{(0)}(t)}, \tag{2.15}$$

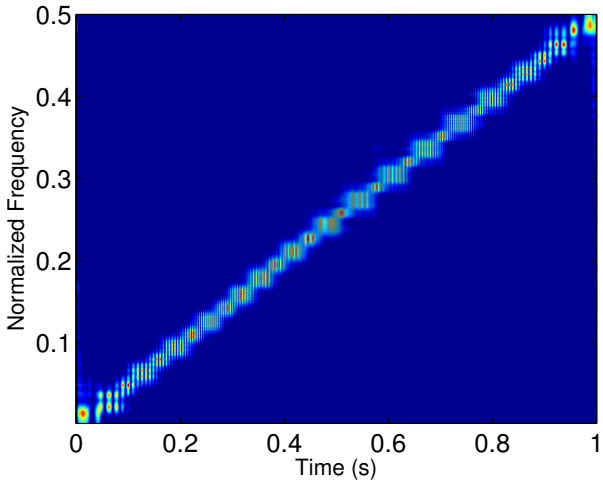


Figure 2.9: (Adaptive MP of a chirp signal with sampling frequency of 1000 Hz and frequency increasing linearly from 0 to 500 Hz.

where, $p^{(0)}(t)$ is the time marginal of $\mathbf{V}^{(0)}(t,f)$. At this stage, $\mathbf{V}^{(1)}(t,f)$ has the correct time marginal, but not correct frequency marginal.

2. In this stage, the frequency marginal is satisfied by multiplying and then dividing the TFD by the desired and the current frequency marginals:

$$\mathbf{V}^{(2)}(t,f) = \mathbf{V}^{(1)}(t,f)\frac{SM(f)}{p^{(0)}(f)}, \tag{2.16}$$

where, $p^{(0)}(f)$ is the frequency marginal of $\mathbf{V}^{(1)}(t,f)$. At this stage $\mathbf{V}^{(2)}(t,f)$ satisfies the frequency marginal condition, but the time marginal could be disrupted.

3. It is shown that repeating the above steps makes the estimated TFD closer to the desirable TF representation of the signal [35]. This follows from the fact that the cross-entropy between the desired TFD and the estimated TFD decreases with the number of iterations.

Fig. 2.9 shows the Adaptive TFD of a chirp signal.

2.2 Selection Criteria for TF Representation Domain

A TFD, $\mathbf{V}(t, f)$, that is non-stationary compatible and could be used for extraction of meaningful features should satisfy the following properties [36]:

- A desirable TF transformation provides a high time and frequency resolution. Therefore, one of the success measures of any TFD lies in how well it can transform the signal on to a TF plane with optimal TF resolution. The ideal case would be to have both time and frequency resolution as high as possible. However, high resolutions in both time and frequency domains cannot exist simultaneously due to the Heisenberg's uncertainty principle. According to Heisenberg's uncertainty principle [37], the TF resolution has to satisfy the condition $\sigma_t \sigma_f >= \frac{1}{2}$, where σ_t and σ_f are the respective time width and frequency width of the TF structure.

- It is invariant to time shift or amplitude scale in the signal. If the structure of the TFD completely changes with a transformation, the TF-based extracted features will also change according to the obtained TFD. Such TFD cannot satisfy the translation invariance property required for the features. Therefore, it is essential for a TFD to follow the same translations as in the signal rather than providing a completely new TF transformation.

- The suitable TF representation provides non-negative TF values:

$$\mathbf{V}(t, f) \geq 0 \tag{2.17}$$

  In order to produce meaningful features, the value of the TFD should be positive at each point; otherwise the extracted features may be very difficult to explain. For example, mean of a negative TFD at a given time might be negative, which means that the instantaneous frequency is also negative. In real-world applications, presence of negative energy or negative instantaneous frequency cannot be interpreted [32].

- satisfies correct time and frequency marginals:

$$\int_{-\infty}^{+\infty} \mathbf{V}(t,f)df = |x(t)|^2, \tag{2.18}$$

$$\int_{-\infty}^{+\infty} \mathbf{V}(t,f)dt = |X(f)|^2, \tag{2.19}$$

where, $\mathbf{V}(t,f)$ is the TFD of signal $x(t)$ with Fourier transform of $X(f)$. The TFD which satisfies the non-negativity and marginal criteria is called positive TFD [31]. A positive TFD with correct marginals estimates a high resolution estimate of the true joint TF distribution of the signal. Such a TFD provides a high TF localization of the signal energy, and it is therefore a suitable TF representations for feature extraction of non-stationary signals.

2.2.1 TF Localization Criteria

Before we move on to the selection of the right TFD for our application, ie. pattern classification, we highlight the importance of the last two criteria mentioned above as related to TF quantification. In this section, we prove that non-negativity (Eqn. 2.17) and marginal criteria (Eqn. 2.19) guarantee the high TF localization of the constructed TFD.

The simplest form of a signal is a one sample signal as is denoted with the following equation:

$$x(t) = A\delta(t-t_0)\text{sin}(2\pi f_0 t), \tag{2.20}$$

where A, t_0 and f_0 represent energy, temporal location, and instantaneous frequency of the above single sample, respectively. In this section, we next find the criteria of the TF transformation that guarantees an accurate time and frequency localization of the given discrete signal in the TF plane. Since any signal is composed of several single sample signals as explained in Eqn. 2.20, we can generalize the criteria for TF localization of one sample signal to any given signal. The problem to be proved is explained as follows:

$\Longrightarrow$ For the signal given in Eqn. 2.20, the obtained TFD $\mathbf{V}(t,f)$ provides a high TF localization if the non-negativity and marginal criteria are satisfied.

Proof: According to the marginal criteria in Eqn. 2.19, at each time (t), the following equation is correct:

$$\int_{-\infty}^{+\infty} \mathbf{V}(t,f)df = |A\delta(t-t_0)\text{sin}(2\pi f_0 t)|^2, \tag{2.21}$$

Because of δ function, the above equation is non-zero only when $t = t_0$:

$$\int_{-\infty}^{+\infty} \mathbf{V}(t,f)df = \begin{cases} 0, & t \neq t_0 \\ |A\text{sin}(2\pi f_0 t_0)|^2, & t = t_0 \end{cases} \tag{2.22}$$

If sum of a set of non-negative numbers is zero, we can conclude that all the numbers in that set are also zero. With the same logic and considering the fact that $\mathbf{V}(t,f)$ is non-negative, we conclude that the value of $\mathbf{V}(t,f)$ at any time or frequency has to be zero except at time equal to t_0. Therefore, Eqn. 2.22 can be re-written as below:

$$\begin{array}{l} \mathbf{V}(t,f) = 0, \quad t \neq t_0 \\ \int_{-\infty}^{+\infty} \mathbf{V}(t,f)df = |A\text{sin}(2\pi f_0 t_0)|^2, \quad t = t_0 \end{array} \tag{2.23}$$

Repeating the above procedure for the spectral marginal, we obtain the following equations:

$$\begin{array}{l} \mathbf{V}(t,f) = 0, \quad f \neq f_0 \\ \int_{-\infty}^{+\infty} \mathbf{V}(t,f)dt = |A\text{sin}(2\pi f_0 t_0)|^2, \quad f = f_0 \end{array} \tag{2.24}$$

where $A\text{sin}(2\pi f_0 t_0)$ is the Fourier transform ($\mathcal{FT}$) of the signal $x(t)$ in Eqn. 2.20:

$$A\delta(t - t_0)\text{sin}(2\pi f_0 t) \overset{\mathcal{FT}}{\Longleftrightarrow} A\text{sin}(2\pi f_0 t_0) \tag{2.25}$$

Combining Eqns. 2.23 and 2.24, it can be seen that the TFD, $\mathbf{V}(t,f)$, is zero at all the time and frequency points, except at t_0 and f_0, which means that $\mathbf{V}(t,f)$ is a two dimensional Dirac's delta function in TFD. According to Eqns. 2.23 and 2.24, the value of this direct function at (t_0, f_0) is equal to $|A\text{sin}(2\pi f_0 t_0)|^2$, or $\mathbf{V}(t,f) = \delta(t - t_0, f - f_0)\,|A\text{sin}(2\pi f t)|^2$.

We showed that if the TFD of one sample signal is non-negative and satisfies the time and spectral marginals, it provides an accurate TF representation of a single sample signal; i.e., the constructed TFD is a true distribution of the signal energy and provides a correct TF localization of the signal. The non-negativity and true marginal rule can be extended to any discrete signal.

2.3 Critical Review of TFD Methods

As mentioned in this chapter, several TFD methods exist; however, not all the methods are non-stationary compatible, or are suitable TF representations for non-stationary feature extraction purposes. This section performs an analytical comparison among the TFDs explained above.

2.3.1 TFD Illustration

A non-stationary signal is constructed as the sum of two linear frequency modulated signals (chirps) with the equations given below:

$$x(t) = sin(2\pi f_1 t + c_1 t^2) + sin(2\pi f_2 t + c_2 t^2), \tag{2.26}$$

where f_1, f_2, c_1 and c_2 are the coefficients of the chirps. One second of signal $x(t)$ is displayed in Fig. 2.10(a). In Eqn. 2.26, the coefficients (f_1, c_1, f_2, and c_2) are selected to obtain two intersecting chirps; one ranging from the frequency of 100 Hz to 400 Hz, and the other one starting from 400 Hz and ending at 100 Hz. Such a signal that is composed of more than one component is called a multi-component signal.

Based on the signal's structure, we expect to evidence two intersecting lines in the constructed TFD, each representing one of the components. Fig. 2.10 displays the TFDs of the chirps using different TF analysis methods. It can be observed from Fig. 2.10(b) that WVD provides a very high TF resolution of the chirps; however, the diamond shape energy observed at the center of the WVD represents energy distributions which we know that do not belong to the chirp signals. These artifacts are due to cross-terms of the two chirps, and damage the efficiency and accuracy of the TF representation. Fig. 2.10(c) displays the spectrogram TF representation of the signal. In the spectrogram TFD, the two chirps are recognizable; however, because of the limited TF resolution of the spectrogram, the area of the rectangles seen in the spectrogram plot are limited by the Heisenbergs TF uncertainty. Hence, the time or frequency localizations of the chirps are rather poor. Scalogram TFD is shown in Fig. 2.10(d). Compared to the previous TFDs, the scalogram offers the least TF resolution. Figs. 2.10(e) and 2.10(f) show MP-TFD and Adaptive-TFD, respectively. Both the distributions provide a high resolution and cross-term free distribution of the signal. In MP-TFD, slight artifacts exist around narrow TF atoms which are removed in Adaptive-TFD.

To further compare the TF analysis methods, the MP-TFD and spectrogram of a speech signal is shown in Fig. 2.11(b). The TFD is constructed using 1000 TF Gabor functions ($I = 1000$). We also obtained the Spectrogram of the same speech sample as shown in Fig. 2.11(c). Comparing

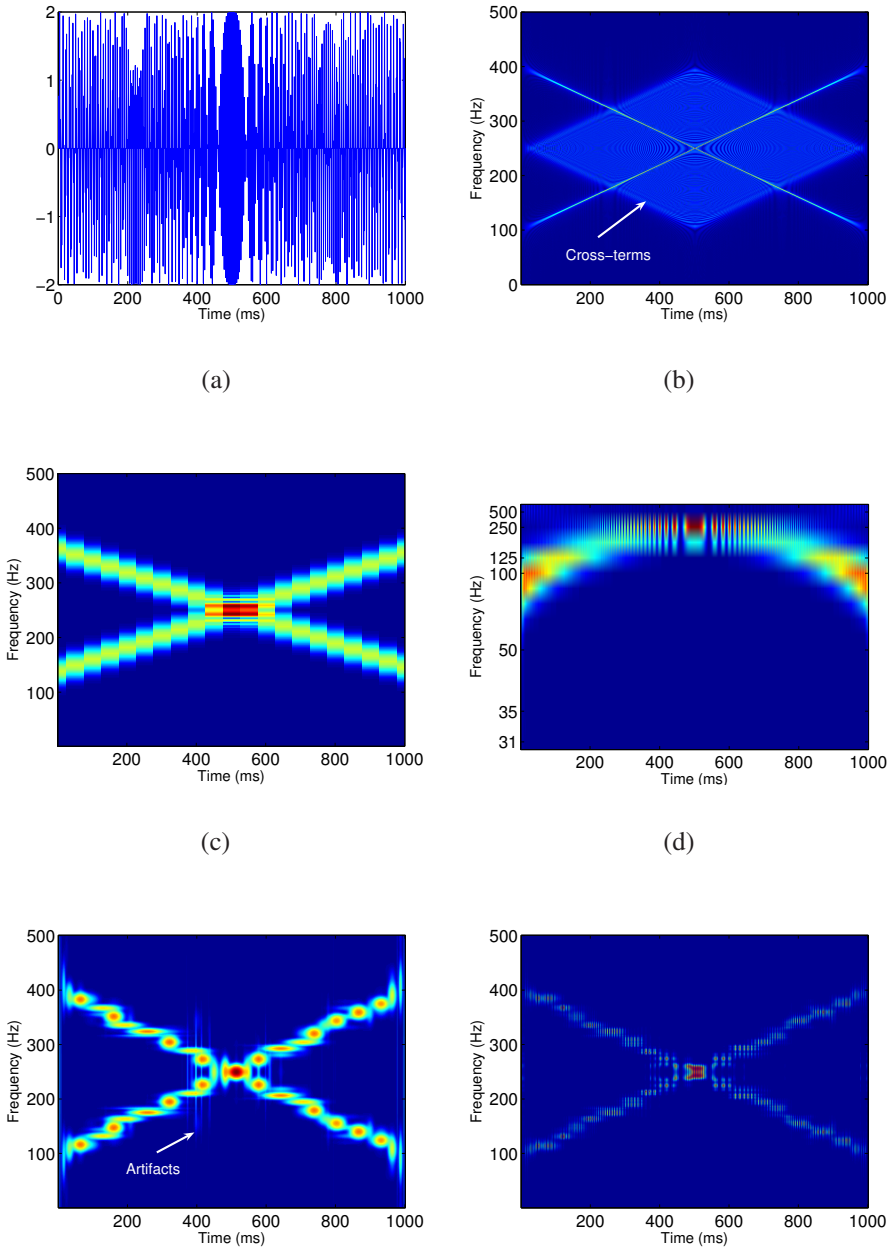


Figure 2.10: Illustration of different TF representations. (a) The signal in temporal plane. (b) The WVD with number of frequency bins equal to the signal length is plotted in logarithmic plane. (c) Spectrogram with FFT size of 1024 points and Kaiser window with parameter of five, length of 256 samples and 220 samples overlap. (d) Wavelet scalogram with complex Gaussian wavelets and 16 scales. (e) MP-TFD with Gaussian atoms and 100 decompositions. (f) Adaptive-TFD using 5 iterations of MCE optimization.

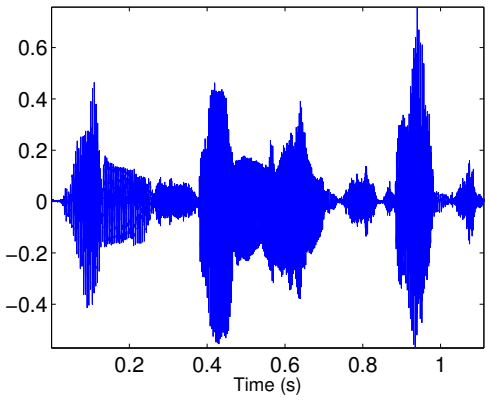
the MP-TFD and Spectrogram of the speech signal, we can observe that MP-TFD presents the TF structure of the uttered speech with high TF resolution compared to Spectrogram. Additionally, the MP-TFD accurately tracks the non-stationarity structure in the speech signal without any cross-terms to damage the TF resolution.

2.3.2 TFD Selection

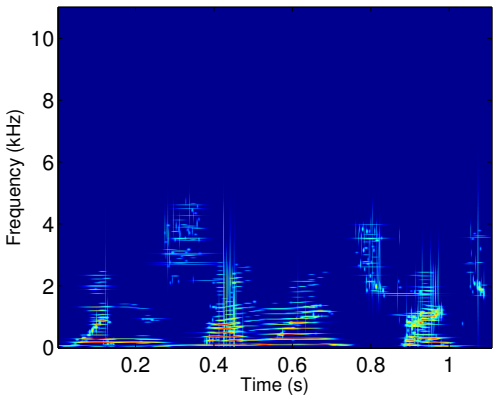
WVD offers high TF resolution; however, the WVD contains interference terms (IT). These cross-terms do not belong to the signal and their presence will lead to incorrect interpretation of the signal properties. This drawback of the WVD is the motivation for introducing other TFDs such as pseudo Wigner-Ville distribution (PWVD), smoothed pseudo Wigner-Ville distribution (SPWVD), Choi-Williams distribution (CWD) and Cohen kernel distribution to define a kernel in ambiguity domain that can eliminate cross terms. These distributions belong to a general class called the Cohen's class of bilinear TF representation [33]. These TFDs satisfy time and frequency marginals; however, the distributions do not always satisfy the non negativity constraint. Therefore the extracted WVD features may not be always meaningful. For example, in WVD, the expectation value of the square of the frequency at a fixed time can become negative, which does not make any sense [31].

Cohen-Posch TFD, or positive TFD (PTFD), produces non-negative TFD of a signal that does not contain any cross terms. Even though PTFD successfully constructs a positive and high resolution TFD of a given signal, this method cannot be implemented in most cases. In order to calculate positive kernels, the method requires the signal equation which is not usually known. Therefore, although the existence of PTFDs is proven, their derivation process is too complicated to be considered in most of the applications.

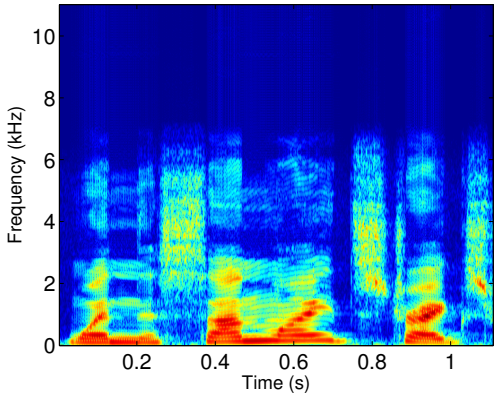
Spectrogram represents a suitable TF representation; however, it suffers from TF resolution trade off; when time is short, frequency resolution is coarse, and vice-versa. To solve the TF resolution requirements wavelets scalogram [37] is introduced. Unlike spectrogram where the time width of the window function is fixed, wavelet scalogram has an adaptive varying time width defined by the scaling parameter. Additionally, scalogram provides a positive and cross-term free TF representation. However, the main drawback of the scalogram is its poor temporal resolution at



(a) Speech sample



(b) MP-TFD



(c) Spectrogram

Figure 2.11: (a) The speech signal is '*When the sun light strikes*' which is spoken by a female speaker and is recorded with 22050 Hz sampling frequency. (b) MP-TFD of the speech sample is calculated with Gabor dictionary and 1000 iteration. (c) Spectrogram is calculated with FFT size of 1024 points, and Kaiser window with parameter of 5, length of 256 samples and 220 samples overlap.

low frequency regions of the TF plane and poor spectral resolution at high frequencies. Therefore, scalogram cannot efficiently display TFD of signals containing components with short durations and low frequencies, or vise versa. The other drawback of the classic wavelet scalogarm for feature extraction purposes is its variability to transformations such as time shift or scaling.

MP-TFD overcomes the shortcomings of both wavelet scalogram and spectrogram. Unlike spectrogram that a fixed-length window is applied to the signal, in MP-TFD at each iteration, the algorithm adaptively selects the window length that best suits the signal. Due to the over-complete dictionary, MP-TFD yields a TFD that achieves any adaptive TF resolution at any part of the TF plane satisfying the Heisenberg's condition. The redundancy of the TF dictionary used for TF decomposition of audio signals provides an extreme flexibility to model a signal as accurately as possible. This property of MP-TFD allows to construct the TFD that best approximate the non-stationarity characteristics in audio signals.

Additionally, MP-TFD is positive and cross-term free. As explained earlier in this chapter, although WVD distribution is a powerful TF representation with high TF resolution, when more than one component is present in a signal, the TF resolution will be diluted by cross-terms. However, when WVD is applied to single components, their summation is a cross-term free TFD. Since practical signals are composed of several components, the cross-term free MP-TFD provides a TF representation adaptive to the TF structure of a given signal.

Even though we were successful in achieving varying TF resolutions over TF plane, the constructed MP-TFD did not satisfy temporal or spectral marginals. As demonstrated earlier in this chapter, to obtain a very accurate TF representation with high time and frequency resolution, the conditions shown in Eqn. 2.19 have to be satisfied. Despite MP-TFD, Adaptive-TFD performs an iterative optimization routine to guarantee the marginal criteria. Hence, Adaptive-TFD satisfies all the criteria mentioned in Section 2.2. The properties of different TFD techniques are summarized in shown in Table 2.1.

Table 2.1: Desirable TFD Properties for TF Quantification

Property	Positivity	Marginals	TF Resolution	TF Localization	Practical Ability
WVD		×	×		×
PTFD	×	×	×	×	
Spectrogram	×				×
Scalogram	×				×
MP-TFD	×		×		×
Adaptive TFD	×	×	×	×	×

2.4 Chapter Summary

This chapter presented the comparison of few well known TF distribution techniques from their TF resolution and feature extraction point of view. It is obvious from the above presentation that the adaptive TF transformation based on matching pursuit would be the appropriate TF transformation tool for the proposed work. Adaptive-TFD provides flexible presentation with excellent TF resolution and the TFD generated from it is of high quality with no cross terms. Further to the above contribution, in Section 2.2.1, we proved that non-negativity and marginal properties of a single sample TFD guaranteed that the constructed TFD was a true distribution of the signal energy and provided a correct TF localization of the signal. This property is beneficial in extracting efficient features with high TF representations that can potentially improve the pattern recognition and decision making procedure.

Fig. 2.12 displays the contribution flowchart, and the highlighted block in this figure shows the progress of the work in this chapter. Chapters 3 and 4 focus on TF quantification of known TF features and embedded TF features, respectively. Quantification of unknown TF features is studied throughout Chapters 5, 6, 7, and 8.

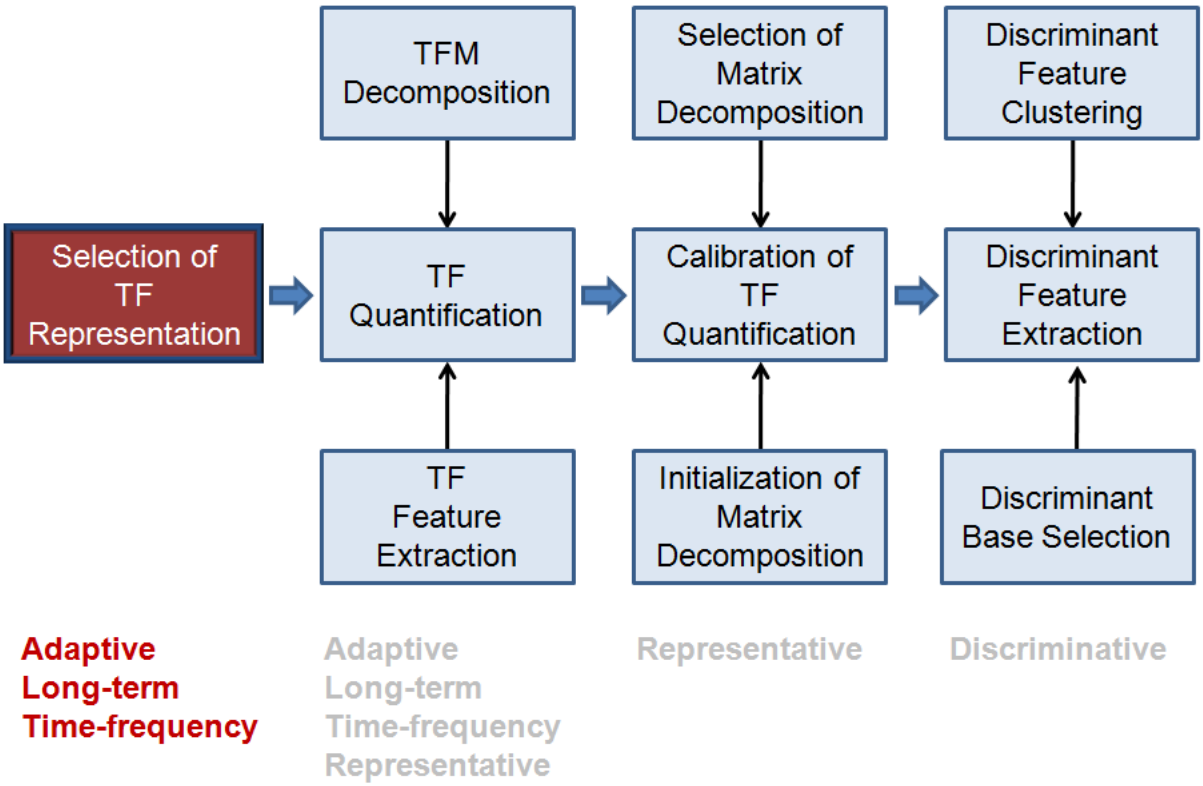


Figure 2.12: Flowchart of the proposed contributions.

Chapter 3

KNOWN TF FEATURE DETECTION

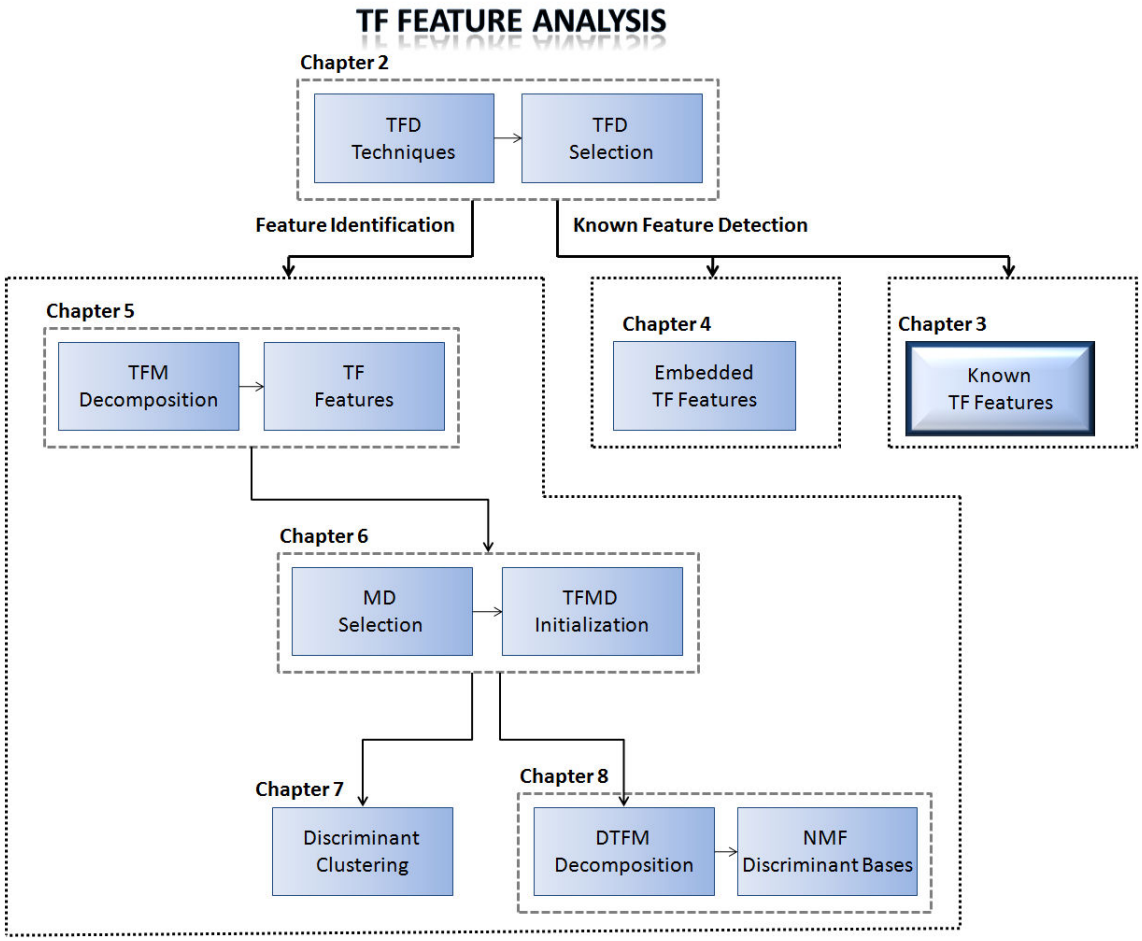


Figure 3.1: Chapter 3 - Known TF quantification and detection.

3.1 Motivation

ONE of the challenges in a pattern detection system is to detect the signatures that make a given signal distinguishable among others. In general, there are two type of features: i) Unknown features: some applications first require to reveal the distinctive characteristics of the signal, and then classify the signal based on the existence or the absence of that pattern in the signal. For example, in a pathological speech detection, before the system can perform any pattern recognition, it has to quantify the signal patterns that cause the abnormality in the speech. ii) Known features: on the other hand, there are situations in which the characteristics of interest are known to us. In such scenarios, our objective is to find a feature extraction technique that quantifies the pattern of interest. There are two categories of known feature detection approaches:

1) Known features: The features of interest are obtained as part of the signals' structure which is known to us. The face detection application is an example of such feature detection prospective, in which the face being searched is the pattern of interest. Another example is the detection of every-other beat variations in the ventricular repolarization portion of an electrocardiogram (ECG) signal as a risk indicator of sudden cardiac death. In both scenarios mentioned above, if the right feature extraction methodology is utilized, the extracted signatures can successfully detect the characteristics of interest.

2) Embedded features: The features are derived from an intentionally embedded external signature. The main application of such deliberately embedded features is in multimedia security purposes. In multimedia security systems, a known signature is added to the host data such that the added message is invisible and secure in the data, but can partly or fully be recovered later on if the correct cryptographically secure key is used.

Having said this, in the present chapter, we focus on quantification of a known structure as part of the signal's structure. Detection of a signal's amplitude is among the most important applications of signal quantification, where the goal is to track the amplitude of the signal at a given frequency. Fig. 3.2(a) displays a simplest of signal with known amplitude structure. This signal is constructed as shown below:

$$m(t) = (-1)^t, \tag{3.1}$$

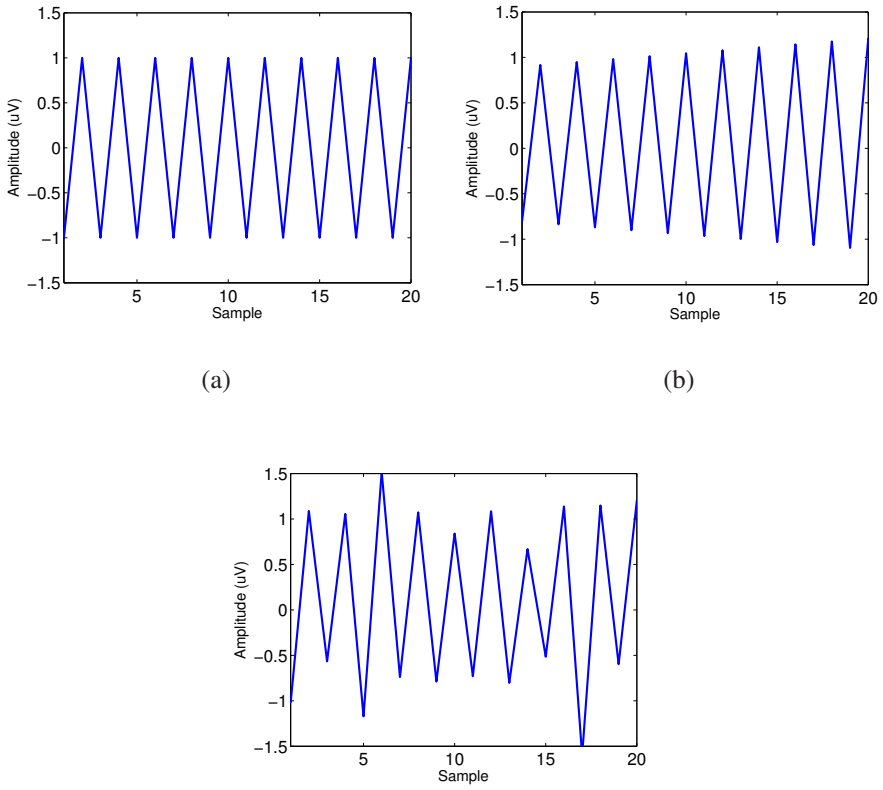


Figure 3.2: (a) A simplest signal with amplitude alternation structure. (b) The same signal with data non-stationarity presented in the amplitude. (c) The noisy version of the signal.

where $m(t)$ is the amplitude of the signal at time sample t. The amplitude of this signal is alternately changing at every other sample. In order to detect such a structure in a given signal, one might suggest to find the amplitude of the signal at consecutive samples, and check if the amplitude alternating structure is evidenced. However, this procedure is not very straight forward in the presence of data non-stationarities or noise as illustrated in the following examples. Fig. 3.2(b) displays the same signal with varying amplitude values. As it can be seen from the plot, the amplitude of the signal is not fixed over time samples, and the alternating pattern is degraded by the non-stationarities in the signal. The noisy version of the same signal with signal-to-noise ratio (SNR) of 10 dB is shown in Fig. 3.2(c). It can be observed that the signal amplitude has been altered due to noise. Despite the existence of the alternating pattern in the signal, the amplitude measurements at the temporal samples cannot identify the alternans.

As demonstrated in the above example, and as our goal to detect the patterns of interest in real-world signals with non-stationarities and presence of noise, in this chapter we focus on developing a robust quantification approach that flexibly tracks the pattern of interest in a given signal. Fig. 3.3 shows the contribution of this chapter. First, we explain two conventional signal quantification approaches. Next, we focus on developing a quantification technique that compensates the limitations of the traditional techniques, and finally we visualize the proposed TF quantification method through a practical application.

3.2 Amplitude Quantification Techniques

3.2.1 Temporal Approach

The simplest approach to quantify the amplitude alternating structure of a signal is a temporal-based technique. This technique calculates a temporal feature over M consecutive samples as shown below:

$$a_T = \sqrt{\frac{\sum_{t=1}^{M} m^2(t)}{M}} \tag{3.2}$$

where a_T is the temporal feature. The above equation calculates the average of the energy at M samples as an estimate of the alternating energy. If the noise added in the signal is Gaussian and

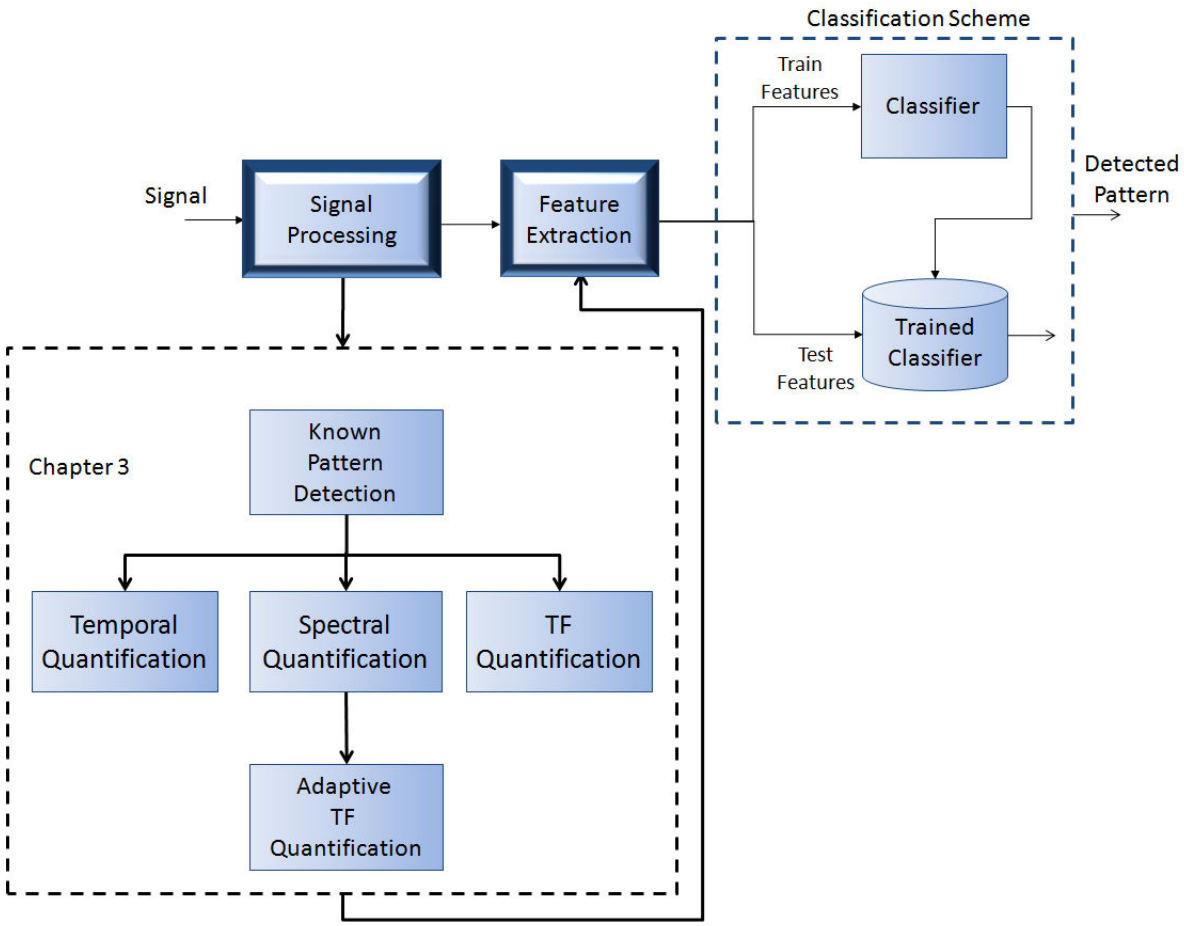


Figure 3.3: Chapter 3 - Known TF Feature detection

the analysis window is long enough, the above procedure produces an acceptable estimate of the signal's amplitude. Using this method, the temporal feature (a_T) is calculated to be 1.1 μV (10 % error) for the noisy signal in Fig. 3.2(c). The extracted feature for the noisy signal is more than 0μV indicating that the alternating pattern is detected in this signal.

3.2.2 Spectral Approach

Another approach is to perform the quantification procedure in spectral domain. The Fourier transform of the original and noisy signal in Fig. 3.2 are displayed in Figs. 3.4(a) and 3.4(b), respectively. As it can be seen in these figures, the alternating amplitude is mapped to the normalized frequency of 0.5. The spectral feature is calculated as follows:

$$a_S = \frac{|\mathcal{FFT}_m(0.5)|}{M} \tag{3.3}$$

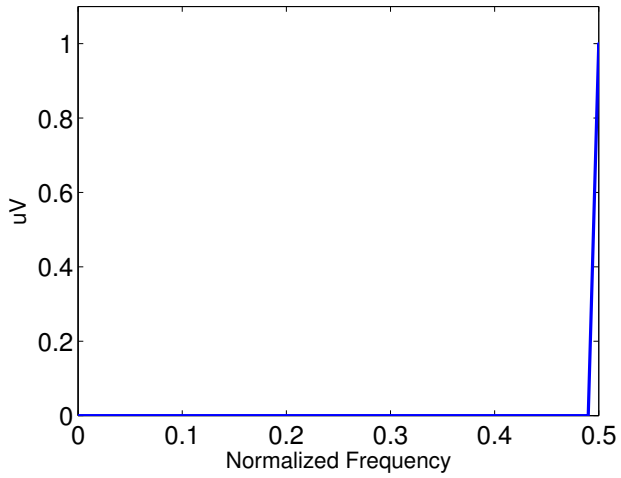
where a_S is the spectral feature, and $\mathcal{FFT}_m(0.5)$ is the magnitude of the Fourier transform of the signal at normalized frequency of 0.5. If the analysis window is large enough, this method accurately extracts the signal's magnitude as the spectral feature. The spectral feature of the noisy signal is 0.98 uV verifying that the method successfully detected the alternating pattern in the noisy signal.

3.2.3 Limitations of Temporal and Spectral Approaches

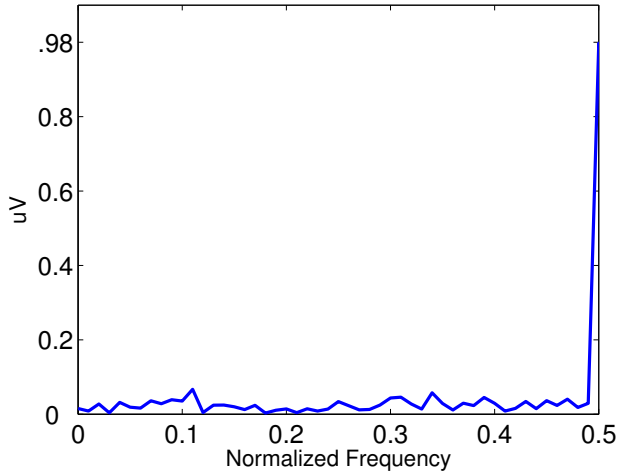
The spectral and temporal approaches assume the stationarity of a signal over M samples:

$$m(t) = a(-1)^t, \text{ for } t = 1, ..., M, \tag{3.4}$$

where a is a fixed magnitude, and m is a vector representing the signal over M samples. Therefore, any changes in magnitude over the M-sample frame will not be accurately tracked by either techniques. This is illustrated in Fig. 3.5 where the amplitude is estimated in a synthetic signal in the presence of magnitude increase from $10\mu V$ at sample 82 to $30\mu V$ at sample 113. In this figure, the change in magnitude over 32 samples is inaccurately represented with both approaches as a change over 64 samples. As shown, abrupt changes and transients in amplitude cannot be picked up immediately using neither the spectral nor the temporal approaches.

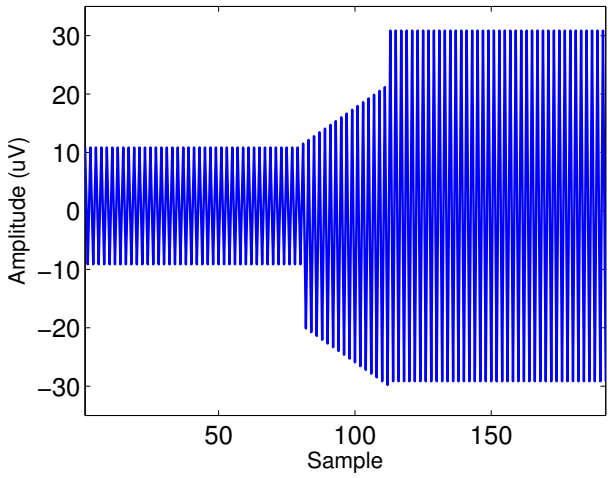


(a)

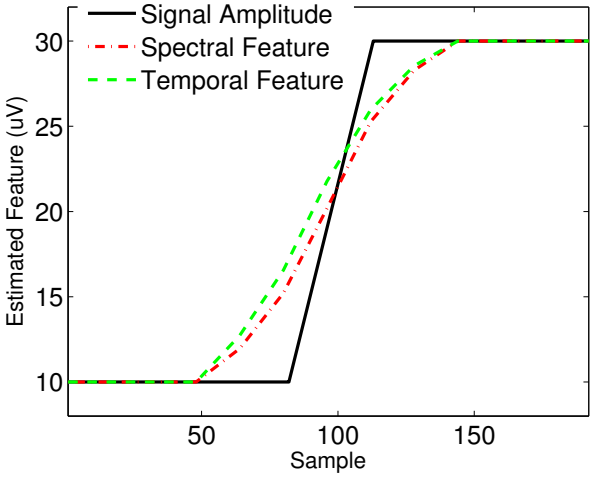


(b)

Figure 3.4: (a) The Fourier transform of the signal in Fig. 3.2(a). (b) The Fourier transform of the signal in Fig. 3.2(c)



(a)



(b)

Figure 3.5: A signal magnitude is measured with spectral and temporal approaches using a 64-sample analysis window, which is shifted by 16 samples in order to track changes in magnitude over the entire 336-sample signal. The solid line depicts an increase in amplitude linearly from 10 to 30 μV from sample 226 to 257. Note the inaccuracy in amplitude measurement using the spectral and temporal approach.

Although a shorter analysis window may permit better magnitude tracking, both the spectral and temporal techniques fail to detect the alternating pattern when only few samples are considered in the analysis. In terms of the spectral technique, a shorter window results in a low resolution Fourier plot with a smaller number of frequency bins, and as a result it will underestimate the alternating pattern. In terms of the temporal approach, if enough samples are not considered in Eqn. 3.2, the averaging cannot eliminate the effect of noise, hence the method will overestimate the value of the alternating amplitude.

An ECG simulation study recently showed that the temporal technique was less robust than the spectral method in detecting T wave alternans (TWA) in the presence of noise contamination [38]. On the other hand, the temporal averaging approach had the potential to provide more accurate signal quantification under data non-stationarity. In this chapter we tackle these limitations in the temporal and spectral techniques, and aim to develop a new approach that is not only flexible to the amplitude non-stationarities, but also robust to noise and outliers.

3.3 Proposed Adaptive TF Quantification

As explained above, the spectral approach is robust to noise, but the issue with this technique was its limitation to track the non-stationarities. In this study, our approach is to remove the stationarity assumption over time, and focus our attention on developing a technique that tracks the non-stationarities over the analysis window. In our new approach, we describe signals by re-writing Eqn. 3.4 as follows:

$$m(t) = a(t)(-1)^t \text{ for } t = 1, ..., M, \tag{3.5}$$

where $a(t)$ represents the amplitude of each time t that may vary over M samples. Fourier transform can no longer be applied to such a time varying structure. In Chapter 2, we explained Adaptive TFD which adaptively tracked the signal's non-stationarity structure. Therefore, instead of the Fourier transform, we propose to use the Adaptive TFD. For a given signal ($m(t)$), Adaptive TFD provides a positive and cross-term free TF representation ($\mathbf{V}(t, f)$) that preserves the time

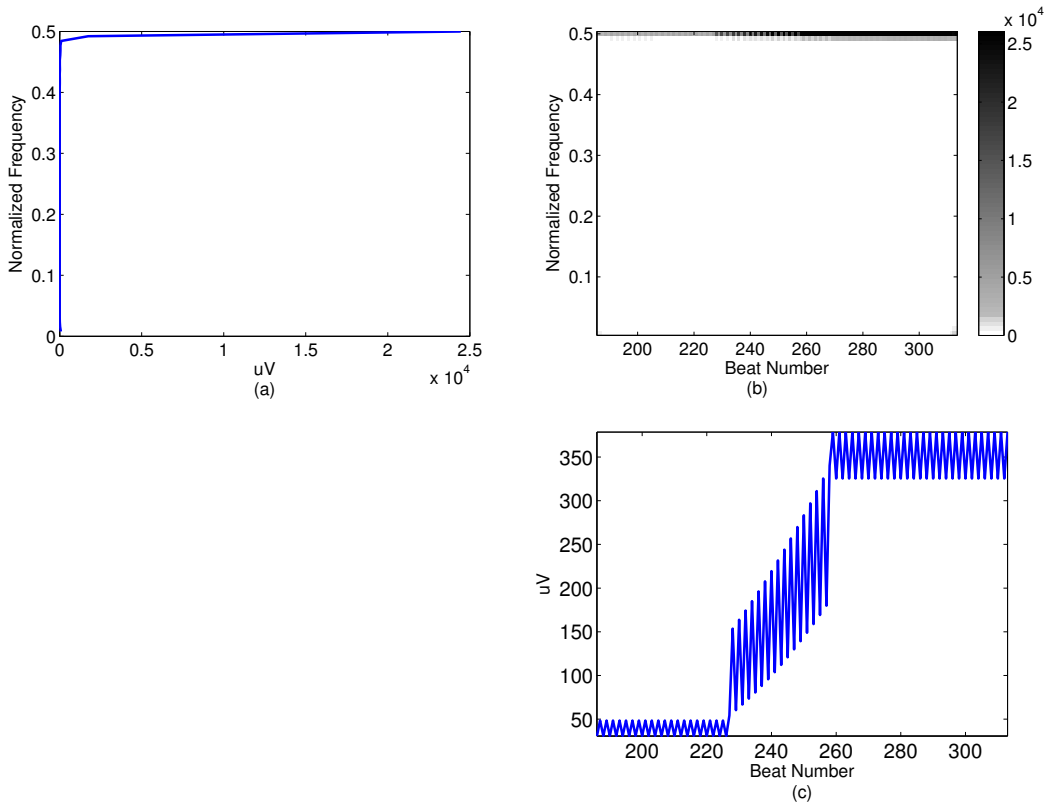


Figure 3.6: Adaptive TFD of the signal shown in Fig. 3.5 (from samples 186 to 313) is constructed. (a) Frequency marginal. (b) Adaptive TFD. (c) Time marginal. As signal's magnitude increases linearly from 10 to 30 μV from samples 226 to 257, the TFD energy (shown in (b)) at normalized frequency of 0.5 also increases as indicated by the color bar. Frequency and time marginals are shown in order to increase the visibility of the changes in TFD.

and frequency marginals as shown below [34]:

$$\sum_{f=1}^{M/2} \mathbf{V}(t,f) = |m(t)|^2 ,$$
$$\sum_{t=1}^{M} \mathbf{V}(t,f) = |\mathcal{FFT}_m(f)|^2 , \tag{3.6}$$

where, $\mathbf{V}(t,f)$ is the Adaptive TFD of signal $m(t)$ with Fourier transform of $\mathcal{FFT}_m(f)$, M is the number of samples in the analysis window and the Fourier transformation. Fig. 3.6 shows the constructed Adaptive TFD of the signal shown in Fig. 3.5 from samples 186 to 313. In Fig. 3.6 (b), the horizontal and vertical axes indicate the time samples and the normalized frequency, respectively. The third dimension, represented by the color at each point, indicates the energy of the signal. The color of the TFD at normalized-frequency of 0.5 changes from light gray to dark

gray, which corresponds to an increase in TFD energy coincident with the rise in the magnitude. The spectral and temporal marginals in Fig. 3.6 (a) and (c), respectively, are shown to better visualize the energy changes in the TFD.

Our proposed Adaptive TF quantification approach uses the constructed TFD above to calculate TF features as follows:

$$a_{AS}(t) = \frac{|\mathbf{V}(t, 0.5)|}{M} \tag{3.7}$$

where $a_{AS}(t)$ is the TF feature at each temporal sample. Fig. 3.7 shows the TF feature extracted from the signal in Fig. 3.2(a). It can be observed that unlike the spectral and temporal approaches, the proposed technique accurately represented the change in magnitude over 32 samples.

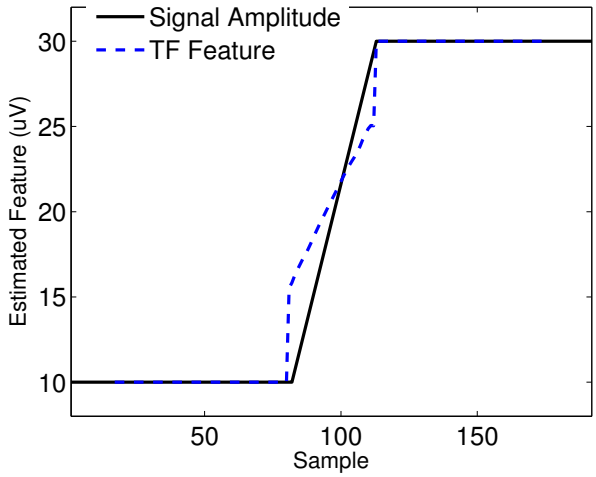


Figure 3.7: The magnitude of the signal in Fig. 3.5(a) is measured using the Adaptive TF quantification technique. The solid line depicts an increase in amplitude linearly from 10 to 30 μV from sample 226 to 257.

3.4 Analytical Comparison of Adaptive TF Quantification with Spectral Approach

We refer to signal magnitude quantified by spectral and Adaptive TF techniques as a_S and a_{AS}, respectively. For comparability with the spectral approach, a_{AS} is defined as the average of the estimated magnitude $a(t)$ over M samples:

$$a_{AS} = \frac{1}{M}\sum_{t=1}^{M} \text{Real}\left\{\sqrt{T(t)}\right\} \tag{3.8}$$

where $T(t)$ is the signal energy at normalized frequency of 0.5. a_S can be derived using Adaptive TF quantification approach as described below:

The constructed TFD satisfies frequency marginals (Eqns. 3.6). Therefore, at each point in TF domain, the sum of the energy over the temporal axis can be written as follows:

$$\sum_{t=1}^{M} \mathbf{V}(t,f) = |\mathcal{FFT}_m(f)|^2 \tag{3.9}$$

In this equation, $\mathcal{FFT}_m(f)$ is the Fourier Transform of the signal, $m(t)$ in Eqn. 3.4:

$$\mathcal{FFT}_m(f) = \sum_{t=1}^{M} m(t)e^{-j2\pi f\frac{t}{M}} \tag{3.10}$$

At normalized frequency of 0.5 ($f = M/2$), Eqn. 3.9 can be written as following:

$$\begin{aligned} \textstyle\sum_{t=1}^{M} T(t) &= |\mathcal{FFT}_m(M/2)|^2 \\ &= \left|\textstyle\sum_{t=1}^{M} m(t)e^{-j\pi t}\right|^2 \\ &= \left|\textstyle\sum_{t=1}^{M} a(-1)^t(-1)^t\right|^2 \\ &= M^2a^2 \end{aligned} \tag{3.11}$$

From the above equation, it follows that

$$a = \frac{1}{M}\text{Real}\left\{\sqrt{\sum_{t=1}^{M} T(t)}\right\} \tag{3.12}$$

and finally, using the constructed Adaptive TFD a_S can be estimated as follows:

$$a_S = \frac{1}{M}\text{Real}\left\{\sqrt{\sum_{t=1}^{M} (T(t))}\right\} \tag{3.13}$$

Comparing Eqns. 3.8 and 3.13, we observe that the spectral approach derives the average magnitude over all M samples, and therefore may underestimate the magnitude in the presence of noise. In contrast the Adaptive TF techniques calculates the average magnitude in the samples where noise is not pervasive [1] which should improve the quantification and detection accuracy of alternating patterns.

[1] If the noise energy at sample t of a segment ($\mu_{noise}(t)$) is more than the energy at normalized-frequency of 0.5 ($T(t)$), the magnitude estimate for that sample will be zero, and will not affect the amplitude estimate for the whole segment.

3.5 Experiment: Electrocardiagram Data Analysis

In this section, we perform a set of experiments to verify the accuracy of the developed feature extraction technique in known pattern detection applications. ECG data analysis is considered as a highly non-stationary and noisy biomedical data.

3.5.1 Background

Each year between 0.5 to 1 million North Americans and Europeans die from sudden cardiac death (SCD) caused by ventricular arrhythmias (VA). However, identifying those patients at risk of SCD remains a formidable challenge as many people are asymptomatic until the VA event occurs, and the majority do not survive the first episode. The standard method for assessing whether a patient is at risk for SCD has been an Electrophysiology (EP) study from inside the heart. However, the EP study is invasive, expensive, and entails some risk to the patient. Therefore, there is a strong need to develop a technology that is quick, noninvasive, relatively inexpensive, yet accurate in identifying those who are at high risk of VA, and benefit from the expensive therapy.

T wave alternans (TWA) has been associated with ventricular arrhythmias. Hence, TWA detection can risk stratify patients with heart disease who may experience sudden death from ventricular arrhythmias. TWA, also called repolarization alternans, is a heart rate dependent phenomenon that manifests on the surface electrocardiogram (ECG) as a change in the shape or amplitude of the T wave every second heart beat. The first cases of visible TWA were reported at the beginning of the 20th century, but it was not until the 1980s, when non-visible (microvolt-level) TWA was measured with the aid of a computer [39]. Since then, TWA is emerging as an important non-invasive marker for sudden cardiac death in patients with heart disease. However, quantification of invisible TWA signals in the presence of confounding effect of biological noise, such as movement, respiration, heart rate change, or premature ventricular contraction (PVC), is a challenging task. There have been attempts to quantify this phenomenon. The current TWA quantification techniques can be divided into two categories: temporal and spectral approaches [40]. The temporal approaches have the potential to provide accurate TWA measurements over data non-stationarities; however spectral methods result in a more optimal measurement under noisy conditions [38]. This section aims to

develop a novel joint temporal and spectral TWA quantification technique that has the advantages of both approaches. Such a technique has to be robust to the following physiological noises:

- Random noise created because of environment noise or subjects' movements.
- Periodic noise caused by respiration.
- Heart rate change due to the subjects' activities during the ECG recording.
- TWA variations over time.
- Ectopics also known as premature ventricular contraction (PVC).
- Phase change in which a premature beat with TWA phase reversal is induced.

T Wave Alternans

Our heart is a muscle that circulates blood in our body. The motion of heart is stimulated by an electrical signal which is known as electrocardiogram (ECG). Each beat begins with an electrical signal from the sinoatrial or SA node which is located in the right atrium. This signal causes the atria to contract, and pumps the blood from atria into both ventricles. Fig. 3.8(a) (a) shows one cycle of the electrical activity of the heart over time captured and externally recorded by skin electrodes. This signal, which is called electrocardiography (ECG), is composed of three sections: P wave, QRS complex, and T wave. The P wave is due to this atrial depolarization. The QRS complex is due to ventricular depolarization, and it marks the beginning of ventricular systole. As the signal passes, the hearts ventricles relax. The T wave is due to the ventricular repolarization. The end of the T wave marks the end of ventricular systole electrically and the heart gets ready for the next cycle.

At the start of the 20th century it was believed that the TWA was fairly rare because of how infrequently they were seen on the ECG. Even so, its presence was recognized as being linked to ventricular arrhythmias and sudden death. Later on, it was found that alternans invisible on the ECG were also significant indicators for these conditions. Heartbeats with T waves with virtually identical amplitudes are referred to as having an A pattern whereas when TWA is present, the T

wave patterns that varies from the normal A pattern on every other beat are referred to as having a B pattern. This continual alternation of the A and B pattern is what is characteristic of the TWA. This can be seen in TWA phenomena is demonstrated in Fig. 3.8(b). As shown in Fig. 3.8(c), the average difference between the T waves in pattern A and B is measured as the TWA value.

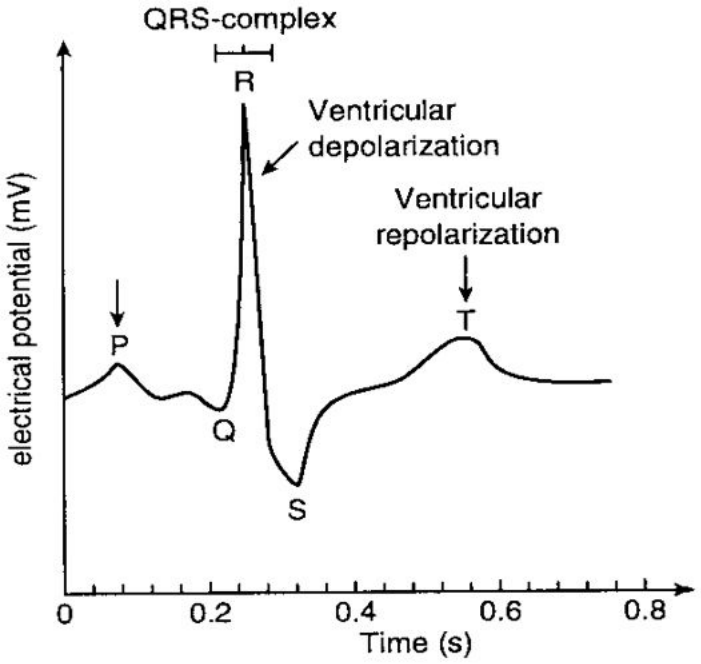
A number of analytical techniques have been proposed to detect microvolt-level TWA from the ECG [40]. These methods can be broadly categorized into: time-domain and the transform-based approach. The time-domain approaches include the correlation method (CM) and the modified moving average method (MMA). The CM detects TWA by computing an alternans correlation index based on a cross correlation technique [41, 42]. The MMA method proposed by Nearing and Verrier [43] computes a beat weighted moving average of odd and even beats, and defines TWA as the difference between the averaged odd and even beats. The spectral method (SM) proposed by Smith et al [39] is an example of a transform-based approach, which uses a periodogram to measure the 0.5 cycle per beat (cpb) TWA frequency component over the aligned T waves. A similar method proposed by Nearing and Verrier [44] known as complex demodulation (CD) fits a sinusoidal signal to the 0.5 cpb frequency of the aligned T waves. Other examples of transform-based approaches are poincare mapping [45] and the periodicity transform method [46].

In the clinical practices, the most commonly used techniques to quantify the TWA signal are SM and MMA.

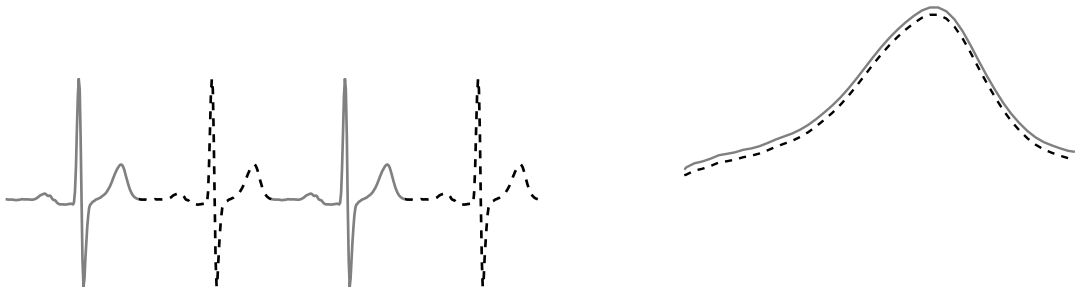
3.5.2 Spectral Method (SM)

The SM transforms a time series of T wave amplitudes across the entire ST segment of consecutive beats to the frequency domain as shown in Fig. 3.9. After pre-processing the ECG recordings, the T wave of each beat is aligned, and matrix $\mathbf{A}_{M\times N}$ is constructed as follows:

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} T_1(1) & T_1(2) & \cdots & T_1(N) \\ T_2(1) & T_2(2) & \cdots & T_2(N) \\ T_3(1) & T_3(2) & \cdots & T_3(N) \\ \vdots & \vdots & \cdots & \vdots \\ T_M(1) & T_M(2) & \cdots & T_M(N) \end{bmatrix} \\ &= [A_1 A_2 \cdots A_N]\end{aligned} \tag{3.14}$$



(a)



(b)

(c)

Figure 3.8: (a) An example of ECG signal. (b) An example a T-wave alternans pattern, where the variations in the T-wave happen every other beat. (c) The difference between successive T waves are called T wave alternans.

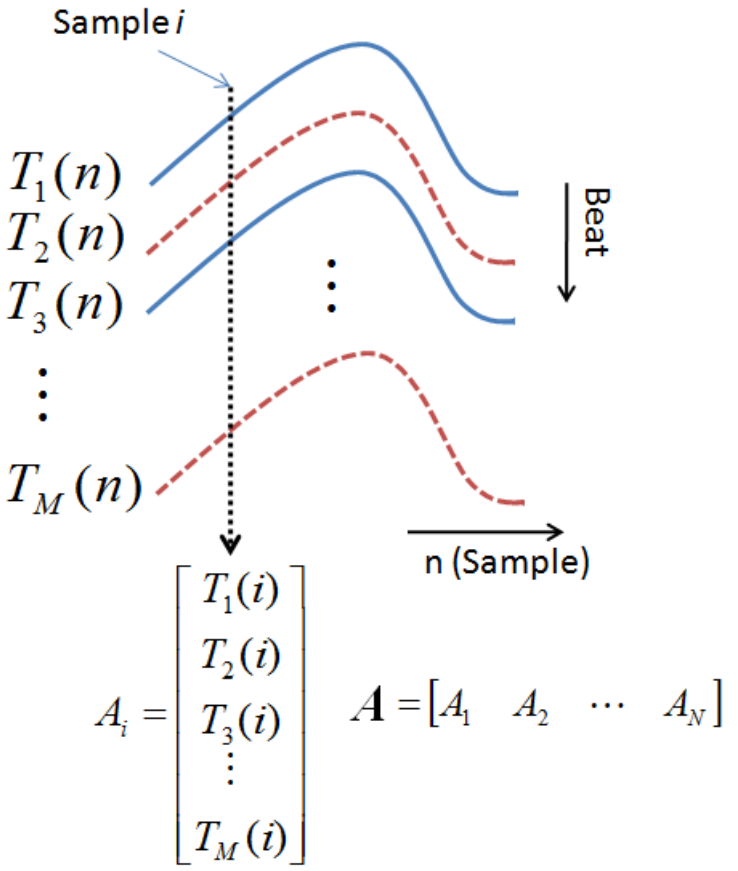


Figure 3.9: Consecutive T waves are aligned, and the T wave amplitude at each sample is transformed into the beat domain.

where $T_j(i)$ (for $j = 1$ to M and $i = 1$ to N) represents the ith sample of the jth T wave, M is the number of heart beats used in the analysis, and N is the length of the T wave. The rows of $\mathbf{A}$ are the T waves of each beat, and the columns show the beat-to-beat variation in T wave amplitude, which is referred to as the beat domain. In the beat domain, there are N signals each with length M, and a sampling rate of 0.5 cycles per beat. According to the definition of TWA, the magnitude of TWA in the beat domain can be measured as the peak of the beat domain signal. This allows measurement of TWA from the spectral magnitude at 0.5 cycles per beat.

3.5.3 Modified Moving Average (MMA)

The MMA denotes A and B beats in an ECG signal with even and odd beats, respectively. In this method, even and odd beats were identified from the ECGs. A median even and a median odd beat were generated and incrementally updated. The incremental updating fraction affects the rate at which the median template tracks changes in T wave magnitude. This fraction was set to 1/16, which corresponds to a 64-beat sampling window for TWA measurement. Using the first odd and even beats to initialize the median odd and median even beats can introduce error in TWA measurement when the ECG recording is noisy. In order to reduce this error, the median odd and median even beats generated by the MMA at the end of the 64-beat sampling window were used to initialize TWA measurement. This modification more closely simulates continuous TWA measurement in ambulatory ECG recordings using MMA.

3.5.4 Adaptive SM

Fig. 3.10 is a schematic diagram of the Adaptive SM. After pre-processing the ECG recordings, the T wave of each beat is aligned, and matrix $\mathbf{A}_{M\times N}$ as shown in Eqn. 3.14. Next, Adaptive TFD is performed to the matrix $\mathbf{A}$, and the average Adaptive TFD for the aligned T waves is constructed as follows:

$$\mathbf{V}_{\frac{M}{2}\times M} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{V}_i \tag{3.15}$$

where $\mathbf{V}_i$ is the Adaptive TF matrix of the ith column in matrix $\mathbf{A}$ (A_i in Eqn. 3.14), M is the number of T waves and N is the length of the T wave. Fig. 3.10 shows the schematic of Adaptive

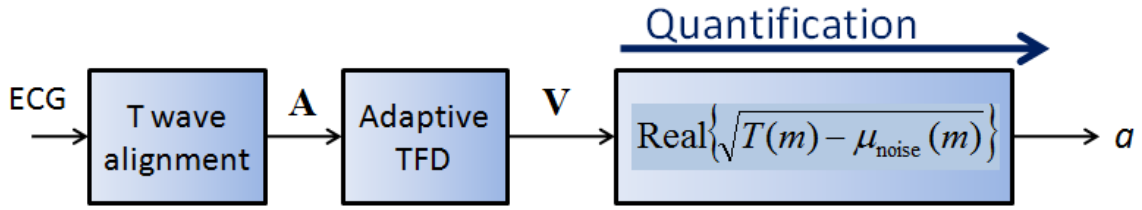


Figure 3.10: Schematic of the Adaptive SM for TWA quantification.

SM, where the constructed TFD is considered to be a series of consecutive T waves (Eqn. 3.5), and it is used to estimate TWA magnitude. The energy of the constructed TFD (**V**) at 0.5 cpb for each beat is considered as the TWA energy at that beat:

$$T(t) = \mathbf{V}(M/2, t) \tag{3.16}$$

As shown in Eqn. 3.6, Adaptive TFD satisfies time marginal. Therefore, combining Eqns. 3.16 and 3.6, TWA at each beat can be derived as follows:

$$a_{ASM}(t) = \text{Real}\left\{\sqrt{T(t)}\right\} \text{ for } t = 1, ..., M, \tag{3.17}$$

As in SM [47], we estimate the noise energy from the energy of the TFD at the spectral bandwidth, 0.44 to 0.49 cpb, and calculate the TWA at each sample with the following equation:

$$a_{ASM}(t) = \text{Real}\left\{\sqrt{T(t) - \mu_{noise}(t)}\right\} \tag{3.18}$$

where $\mu_{noise}(t)$ is the noise energy at beat t. As previously described in [47], similar to SM, a $\text{K}_{score}(i)$ is then calculated as the ratio of the alternans power divided by the standard deviation of the noise in the spectral bandwidth, 0.44 to 0.49 cpb:

$$\text{K}_{score}(t) = \frac{T(t) - \mu_{noise}(t)}{\sigma_{noise}(t)} \tag{3.19}$$

When the K_{score} is larger than 3 [47], the alternans power is greater than the noise level, and the TWA estimation can be considered reliable.

Adaptive SM constructs the average TFD of all the beat series in the aligned T wave and then uses the spectral magnitude of the average TFD at f=0.5 cpb to measure TWA. Since the average

TFD is a matrix with M/2 samples in row and M samples in its column, as seen in Eqn. 3.18, Adaptive SM measures one TWA value for every beat, which enables the method to track beat-to-beat changes in TWA. The tracking capability of the proposed method increases the accuracy of TWA quantification.

3.5.5 Dataset

Synthetic ECG Recordings

Synthetic ECGs were created by the periodic replication of a single QRST complex. The QRST complex is obtained by averaging 10 QRST complexes from ECG lead V4 recorded in a patient during sinus rhythm. Recordings were made at a sampling rate of 1000 Hz, then downsampled to 200 Hz in order to approximate the typical sampling rate of ambulatory ECGs. A schematic of the synthetic ECG signal generator and the TWA analysis is illustrated in Fig. 3.11(a). As shown in this figure, a simulated TWA signal with amplitude a is added to the synthetic ECG. This is achieved by uniformly increasing T wave amplitude of even beats and decreasing T wave amplitude of odd beats across the T wave. An alternative approach could be to use a physiological alternans waveform by multiplying the rectangular TWA by a Hanning window. The use of a rectangular TWA versus a more physiological alternans waveform would not make any difference to the results since alternans is computed at each point on the T wave. Fig. 3.12 compares the performance of the TWA estimation techniques for physiological and uniform TWA magnitude. In Fig. 3.12(a), a Hanning-shape TWA is added across the T waves. The amplitude of the Hanning window at maximum point increases linearly from 1 μV at beat 100 to 15 μV at beat 131, then remains at 15 μV for the next 70 beats, and finally decreases linearly to 2 μV over the next 100 beats. In Fig. 3.12(b), a rectangular TWA uniformly changes the TWA magnitude across the ST interval. As can be seen in these figures, the shape of TWA did not affect the performance of the techniques; however, the use of a uniform TWA magnitude across the ST interval allowed direct comparison between the methods because the SM and Adaptive SM calculate TWA as the square root of mean alternans power across the ST interval, while MMA defines TWA as the mean alternans amplitude across the ST interval.

Ambulatory ECGs are often contaminated by noise from non-periodic Gaussian and periodic noise sources such as muscle (EMG) artifact, and electrode motion artifact due to movement, and respiration. Therefore, Gaussian white noise with an RMS level of α was added to the synthetic ECGs to simulate continuous random noise. In addition, a periodic signal was added to the synthetic ECG with frequency ν cpb to simulate periodic noise. The periodic signal at each frequency was created by half wave rectification of a sine wave with amplitude 25 μV. Real world periodic noise from muscle artifact and electrode motion were obtained from the MIT-BIH Noise Stress Test Database [48]. Muscle artifact can mimic the appearance of ectopic beats and cannot be removed easily by simple filters, as can noise of other types. Data non-stationarity was simulated by changing heart rate linearly from 60 to 100 bpm over 128 beats. Because the T wave duration shortens physiologically at faster heart rates, we applied Bazett's formula [2] to maintain a constant rate corrected T wave duration [49].

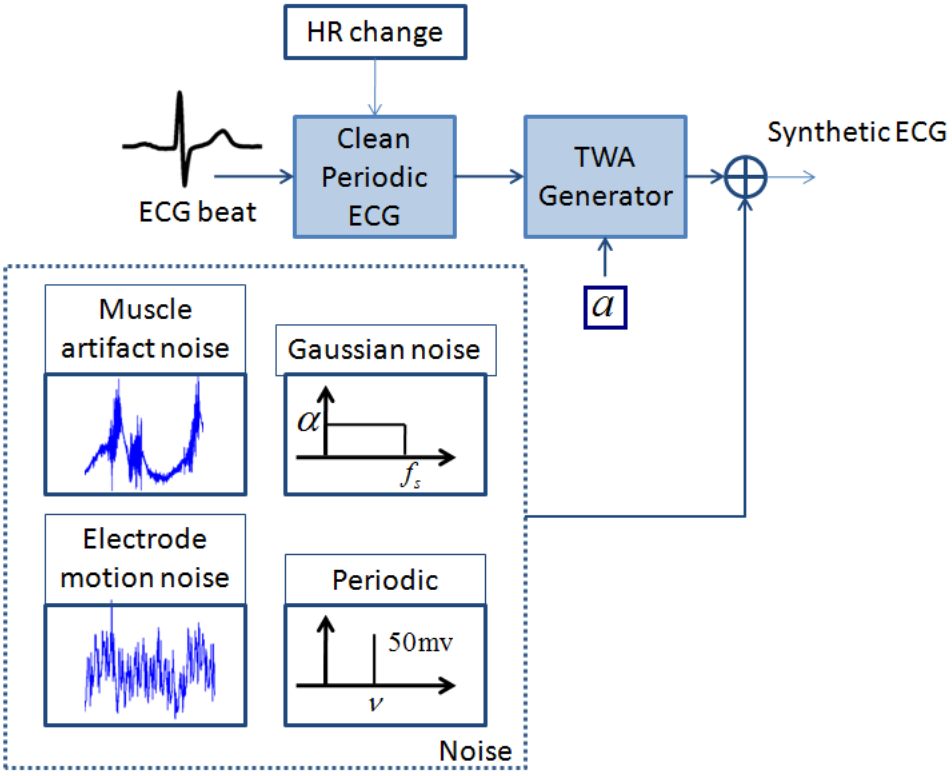
ECG recordings from Invasive Electrophysiology Study

In order to evaluate the effect of data non-stationarity from ectopic beats, the ECG recording from one patient with heart disease was studied. This patient underwent an invasive electrophysiology study for sudden cardiac death risk assessment which involved pacing the right atrium at heart rates of 100, 110 and 120 bpm for 4 minutes each. Standard 12 lead ECG was recorded during the pacing at a sampling rate of 1000 Hz. Recordings from a single ECG lead (lead V2) were used to measure TWA. Baseline wander was estimated by cubic spline interpolation of isoelectric points preceding each QRS onset and was subtracted from the ECG. Each QRS beat was aligned by its QRS onset. Identification of ectopic beats was performed automatically using a QRS morphology matching algorithm and then manually verified. Ectopic beats were replaced with a normal beat derived by averaging the QRST morphology of consecutive normal beats.

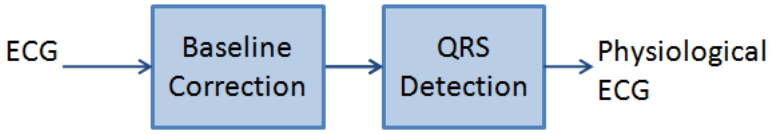
Ambulatory ECG Recordings

Real world ECG recording with inherent noise were obtained from 26 normal subjects who underwent 2 channel ambulatory ECG recording (GE Healthcare, Inc.) for 24-48 hours duration at

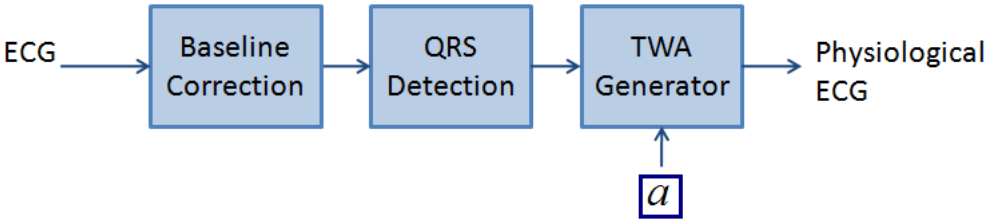
[2] QT / $\sqrt{\text{RR}}$, where QT is the QT interval and RR is the cycle length.



(a) Synthetic ECG dataset



(b) ECG from electrophysiology study recorded in a patient with heart disease.



(c) Ambulatory ECG dataset recorded in normal subjects.

Figure 3.11: Block Diagram of the database generator.

our institution. The ECGs were recorded at a sampling rate of 125 Hz and then exported for custom analysis. Each ECG channel was included as a separate record. Baseline correction and QRS onset annotations were performed as described previously. The noise level of the recordings was determined as the standard deviation (SD) over the first 80 ms of the TP interval after correcting baseline wander. As with the synthetic ECG recordings, a simulated TWA signal of varying amplitudes (a) was added to the ECG. This was achieved by increasing T wave amplitude of even beats and decreasing T wave amplitude of odd beats uniformly across the T wave from a point 40 ms after QRS offset to the end of the T wave. A schematic of these 3 datasets is presented in Fig. 3.11.

3.5.6 Results

We used Adaptive SM to quantify TWA under conditions of non-stationarity, and compared its performance to SM and MMA. After pre-processing the ECG waveform, the T waves of every beat are aligned, and the average Adaptive TFD of the aligned T waves is constructed using adaptive TFD with Gabor atoms, 100 MP iterations and 5 MCE iterations. TWA is measured from the constructed TFD using Eqn. 3.18.

Synthetic ECG Recording

The performances of the three techniques are compared under data non-stationarity and noise.

Simulating Data Non-stationarity

The following non-stationary conditions are simulated: (i) changing TWA magnitude, (ii) changing heart rate, (iii) phase reversal, and (iv) ectopic beats. Fig. 3.12(b) illustrates the TWA signal measured for the synthetic ECG recording using SM, Adaptive SM and MMA. TWA signal non-stationarity is simulated by changing TWA magnitude as shown in the figure. TWA increases linearly from 1 μV at beat 100 to 15 μV at beat 131, then remains at 15 μV for the next 70 beats, and finally decreases linearly to 2 μV over the next 100 beats. Adaptive SM is applied to consecutive 64 beat windows with zero overlap, while the SM and MMA are performed on 64 beat windows that are shifted by 16 beats in order to track changes in TWA magnitude over time. As

evident in this figure, the TWA measured using SM is inaccurate over 32 beats before and 32 beats after a change in TWA magnitude. In contrast, Adaptive SM and MMA accurately track changing TWA magnitude over the same period of time because Adaptive TFD provides an estimate of the TWA signal at each beat.

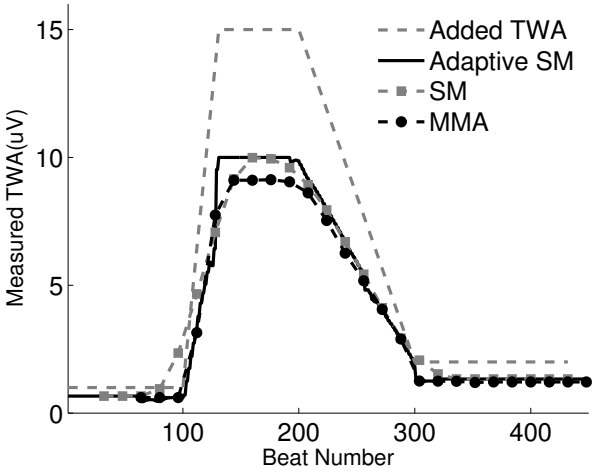
The TWA measured in synthetic ECG during changing heart rate is shown in Fig. 3.13. TWA of 2 μV was added to the synthetic ECG and the measured TWA with SM, Adaptive SM and MMA are computed during heart rate acceleration from 60 to 100 bpm over 128 beats. According to this figure, the accuracy of SM, Adaptive SM and MMA are similar under non-stationary conditions of changing heart rate.

In Fig. 3.14, on beat 128, a premature beat was introduced with TWA phase reversal. As shown in this figure, using the Adaptive SM, the phase reversal resulted in a decline in TWA magnitude over 2 beats, while MMA and SM resulted in a decline within 64 and 48-beat time frames. Fig. 3.15 shows the performance of Adaptive SM, SM and MMA under presence of ectopic beats with no phase reversal. In this figure, the horizontal axis represents the percentages of the ectopic beats inserted in a 500-beat noiseless synthetic ECG, and the vertical axis shows the maximum TWA measured using each technique. As expected, an increase in the number of ectopic beats results in a decline in the maximum TWA measured using all the three techniques; however, under high ectopics, Adaptive SM results in a more accurate TWA measured compared to SM and MMA.

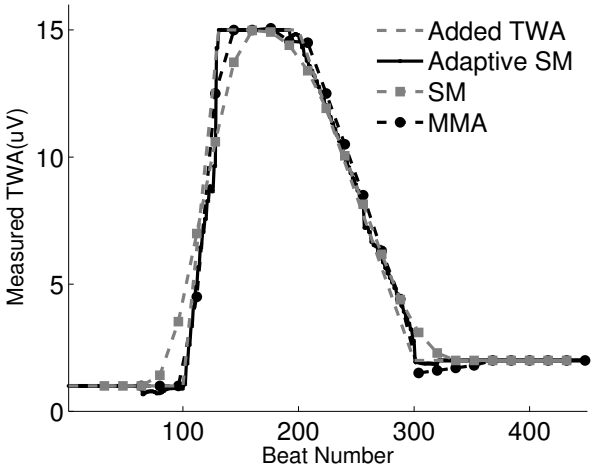
Simulating Noise

The objective of this section is to compare the effect of noise on the accuracy of TWA measurement with Adaptive SM versus SM and MMA. We added simulated TWA of varying magnitude to synthetic ECG with increasing levels of periodic, Gaussian, electrode motion artifact and muscle artifact noise.

The TWA measured from 64-beat synthetic ECGs with added periodic noise (25 μV) is shown in Fig. 3.16. In this figure, TWA measured using SM, Adaptive SM and MMA is plotted as a function of the frequency of the added periodic noise (0.01 to 0.49 cpb). These simulations were performed with either 0 μV or 2 μV simulated TWA signal, the latter being the threshold TWA



(a) Physiological TWA



(b) Rectangular TWA

Figure 3.12: (a) Measured TWA in synthetic ECG using SM, Adaptive SM and MMA using physiological TWA shape. (b) Measured TWA in synthetic ECG using SM, Adaptive SM and MMA using uniform TWA. Comparing (a) and (b), it is concluded that the shape of TWA does not effect the performance of the methods. TWA magnitude increases linearly from 1 μV to 15 μV over 32 beats, then remains constant for 70 beats, and finally decreases to 2 μV over 100 beats. Adaptive and MMA track the TWA changes more accurately compared to SM.

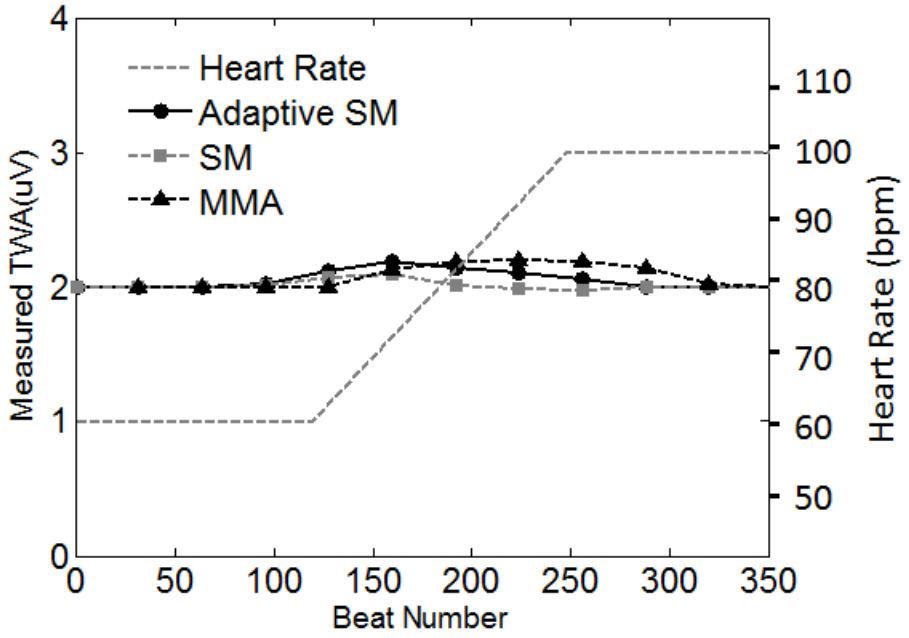


Figure 3.13: TWA measured in synthetic ECG using SM, Adaptive SM and MMA is plotted as a function of increasing heart rate from 60 bpm to 100 bpm over 128 beats. The accuracy of all methods are similar under non-stationary conditions of changing heart rate.

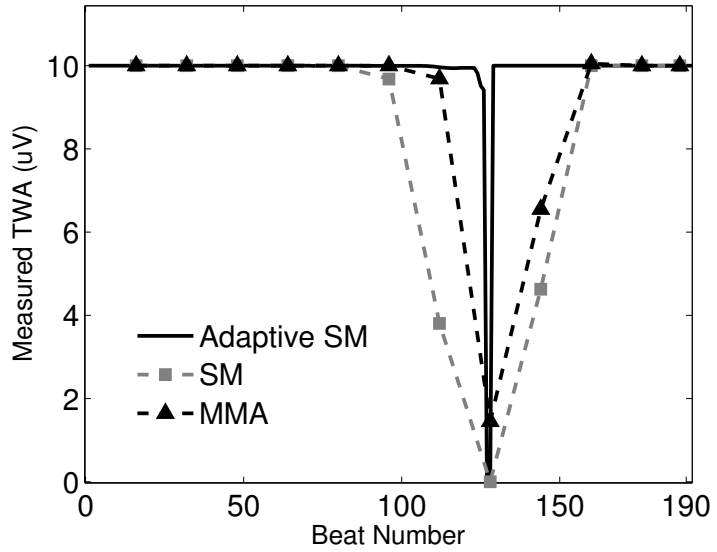


Figure 3.14: TWA measured in a synthetic ECG with a phase reversal at beat 128 using SM, Adaptive SM and MMA. Adaptive SM results in a TWA magnitude decline over a shorter time frame compared to SM and MMA.

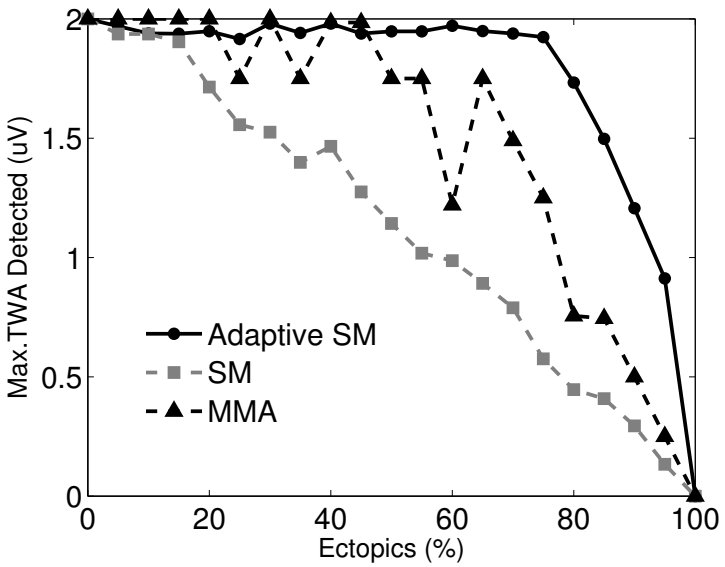


Figure 3.15: Maximum TWA measured in a 500-beat synthetic ECG under different percentages of ectopic beats using Adaptive SM, SM and MMA. Adaptive SM results in a more accurate TWA measurement compared to SM and MMA under the same number of ectopics.

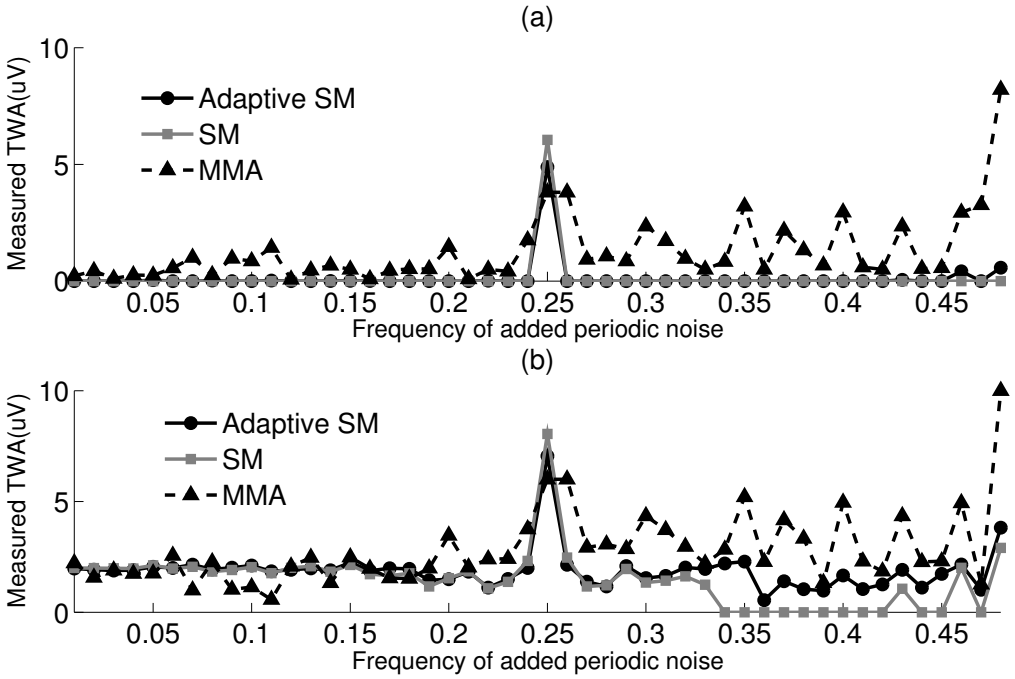


Figure 3.16: TWA measured in synthetic ECG using SM, Adaptive SM and MMA is plotted as a function of the frequency of added periodic noise (0.01 to 0.49 cpb, 25 μV). (a) No added TWA. The accuracy of SM and Adaptive SM is similar without TWA signal, while MMA overestimates TWA. (b) Added TWA of 2 μV. In the presence of 0.26 to 0.49 cpb period noise, Adaptive SM more accurately measures the TWA compared to SM and MMA.

magnitude used in clinical medicine to identify a high-risk patient [47]. Periodic noise with a frequency of 0.25 cpb was detected as TWA using Adaptive SM, SM and MMA. This is a well-known confounder where 0.25 cpb periodic noise produces harmonics at 0.5 cpb [47]. Therefore, the harmonic energy at 0.5cpb is mistaken by TWA energy. The performance of Adaptive SM and SM are similar in the absence of TWA, but in the presence of 2 μV TWA, Adaptive SM more accurately quantifies TWA compared to SM in the presence of periodic noise with frequency 0.35 to 0.45 cpb. With either 0 μV or 2 μV added TWA, TWA was falsely measured using MMA in the presence of 0.26 to 0.49 cpb period noise.

In Fig. 3.17, TWA measurement error (ie measured TWA - added TWA) is compared between SM, Adaptive SM and MMA as a function of increasing RMS noise for added TWA of 2 μV. Each synthetic ECG was analyzed 10 times using different random samples of Gaussian noise with the

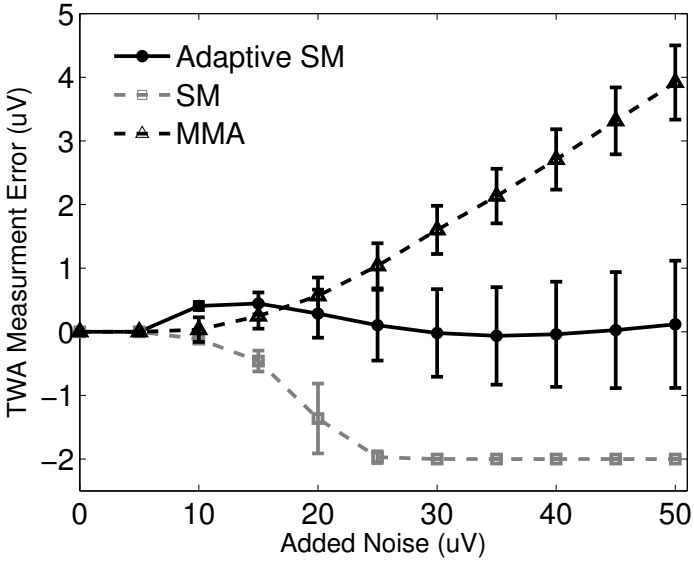


Figure 3.17: TWA measurement error (mean±SD) in synthetic ECGs using SM, Adaptive SM and MMA in the presence of increasing non periodic white Gaussian noise for added TWA of 2 μV, $p < 0.0001$ (SM vs Adaptive SM) for all Gaussian noise values except 20 μV.

same RMS noise level. As shown in Fig. 3.17, when the RMS of the noise is greater than 25 μV, SM does not detect any of the added TWA signal of 2 μV, and MMA overestimates the TWA as the RMS of the noise increases. However, the Adaptive SM measures a TWA of 1 to 3 μV in all the noisy cases. In the presence of noise, Adaptive SM measures noise as the TWA magnitude and it, therefore, overestimates the measured TWA especially when the noise power is high; however, as shown in Fig. 3.17, SM tends to underestimate the measured TWA.

Noise discrimination with SM, Adaptive SM and MMA was also evaluated by adding varying levels of noise to simulated TWA signal of 2 to 14 μV such that the average alternans-to-noise ratio (ANR) increased from -45 dB to -5 dB. For this purpose, the average ANR was defined as below:

$$ANR = \frac{10}{M}\log\left\{\frac{Ma^2}{\sum_{b=1}^{M} u(b)^2}\right\} \tag{3.20}$$

where, M is the total number of beats, a is the added TWA, and $u(b)$ represents the noise value in the b-th T wave. A synthetic ECG signal with 2000 beats (M) was used in this analysis. Figs. 3.18, 3.19 and 3.20 show the contour plots of the TWA measurement error for SM, Adaptive SM and MMA in the presence of Gaussian random noise, electrode motion artifact and muscle artifact

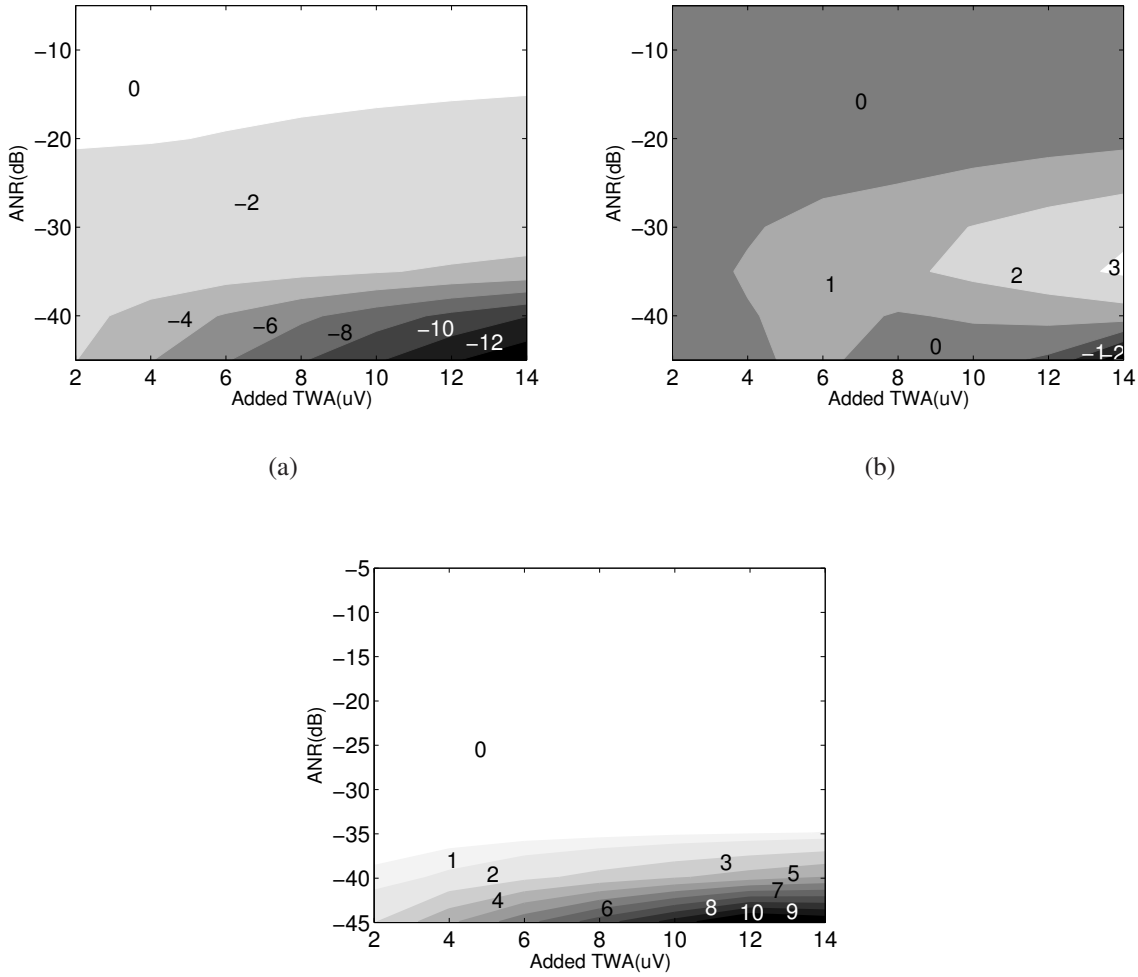
noise, respectively. In Fig. 3.18, when the average ANR decreases, the MMA method and SM significantly overestimate the TWA, while the maximum measurement error with the Adaptive is less that 4 μV in all the noisy cases. In the case of noise from electrode motion artifact and muscle artifact, the contour plots in Figs. 3.19 and 3.20, respectively, show less measurement error with Adaptive SM compared to SM and MMA. As can be seen in Figs. 3.18 (b) and 3.20 (b), for added TWA of 6 to 14 μV, when the average ANR decreases, unlike SM and MMA, Adaptive SM results in a less TWA measurement error. This behavior can be explained with the fact that Adaptive SM is a non-linear approach, and it, therefore, presents a non-linear behavior under noisy condition.

Invasive Electrophysiology Study

The performance of SM and Adaptive SM in the presence of data non-stationarity arising from frequent ectopic beats was evaluated in one patient undergoing invasive electrophysiology testing. Heart rates were controlled with artificial pacing using a quadripolar pacing catheter. Fig. 3.21(a) depicts the heart rate of the patient during the 12 minute recording, and the vertical deflections represents heart rate perturbation from frequent ectopic beats seen predominant at 110 and 120 bpm. The TWA measurement using SM is shown in Fig. 3.21(b) and the shaded areas represent the TWA estimations with K_{score} greater than 3 indicating significant TWA signal over noise. Significant TWA signal is only detected during the first 4 min. at heart rates of 100 bpm when no ectopy is present. SM is unable to detect significant TWA in the presence of ectopy after 4 min.

In the preprocessing stage, these ectopic beats are automatically detected and replaced with a normal beat; however, some ectopic beats fail to be detected such as the one shown in Fig. 3.22 which has a similar morphology to the normal beat. After manually detecting and replacing all ectopic beats, the SM is now able to detect significant TWA at 110 and 120 bpm but there is still some TWA signal dropout as shown in Fig. 3.21(d).

In order to compare the performance of Adaptive SM with SM under the same conditions, we applied Adaptive SM to the ECG recording before and after manual replacement of ectopic beats, and the results are shown in Figs. 3.21(c) and 3.21(e), respectively. Without manually ECG preprocessing, unlike SM, Adaptive SM is able to detect significant TWA signal during most of

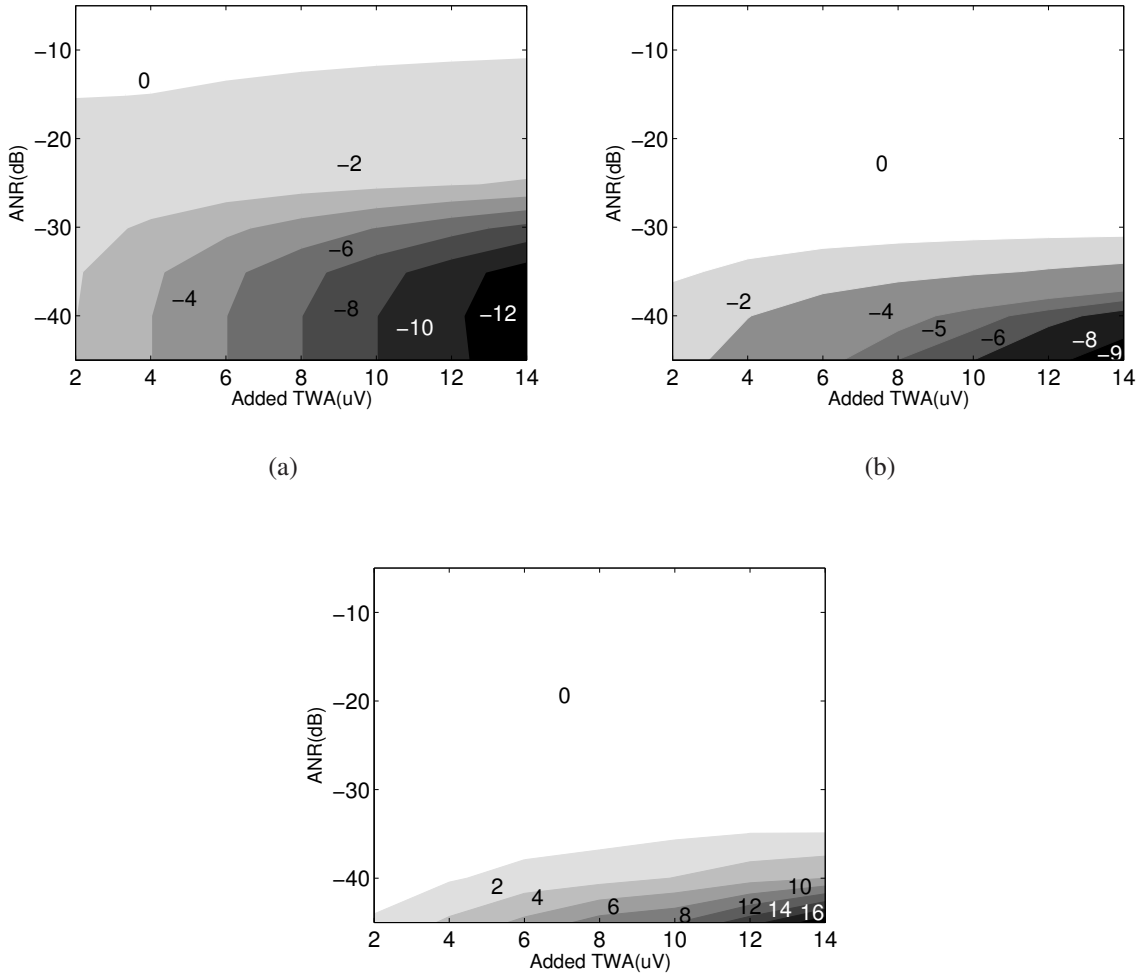


(a)

(b)

(c)

Figure 3.18: TWA measurement error in synthetic ECGs using (a) SM, (b) Adaptive SM and (c) MMA as a function of increasing average ANR and added TWA. Noise was simulated by adding non-periodic Gaussian noise. For the same TWA magnitude and average ANR noise level, the measurement error with Adaptive SM is significantly small compared to SM and MMA.



(a)

(b)

(c)

Figure 3.19: TWA measurement error in synthetic ECGs using (a) SM, (b) Adaptive SM and (c) MMA as a function of increasing average ANR and added TWA. Noise was simulated by adding electrode motion artifact. For the same TWA magnitude and average ANR noise level, the measurement error with Adaptive SM is smaller compared to SM and MMA.

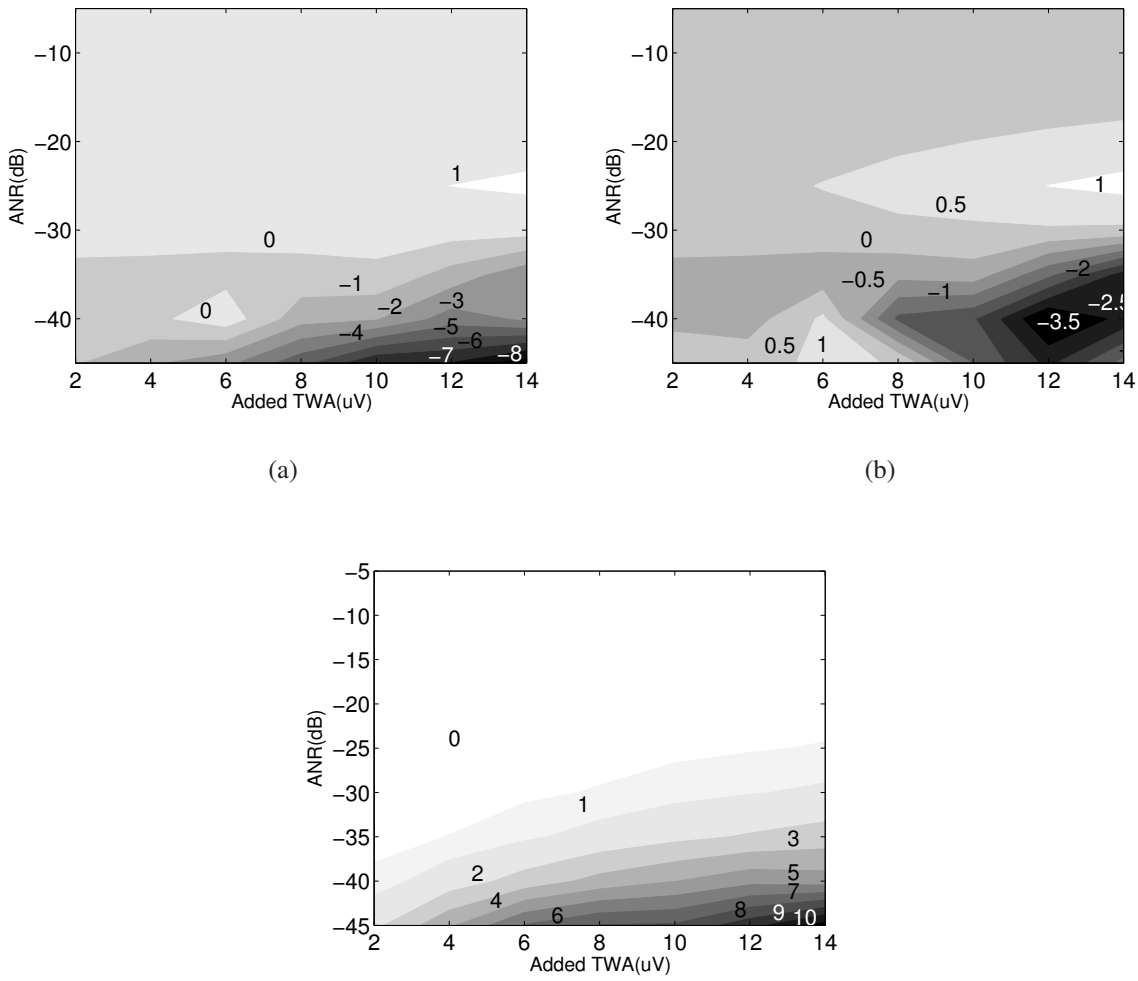


Figure 3.20: TWA measurement error in synthetic ECGs using (a) SM, (b) Adaptive SM and (c) MMA as a function of increasing alternans-to-noise ratio (ANR) and added TWA. Noise was simulated by adding electrode muscle artifact. In low average ANR, the Adaptive SM estimates the TWA more accurately compared to the SM and the MMA method with maximum absolute measurement error of 4 μV compared to 9 μV and 10 μV, respectively.

the 12 min. recording. Thus, Adaptive SM appears to be more robust in TWA signal detection in the presence of frequent ectopic beats compared to SM.

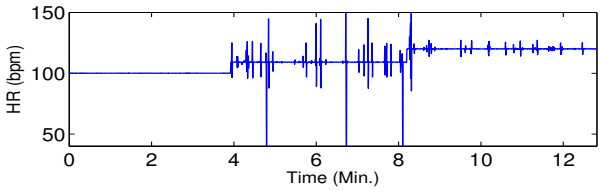
Ambulatory ECG Recording

The ambulatory ECG dataset provided real world recordings from normal subjects with physiological noise and no inherent TWA signal. The mean heart rate in these recordings was 78$\pm$17 bpm and the mean noise level was 40$\pm$67 μV. Simulated TWA signal was added to these recordings with magnitude ranging from 0 to 14 μV. TWA was then measured using Adaptive SM, SM and MMA with 64 beat analysis windows applied to 500 consecutive beats of each ECG recording channel. In Fig. 3.23, the TWA measurement error is compared between the two methods as a function of increasing TWA magnitude. As can be seen in this figure, the MMA technique falsely measures the physiological noise in the ambulatory ECG recordings as TWA, while Adaptive SM and SM accurately measure the TWA magnitude added to the ECG recording.

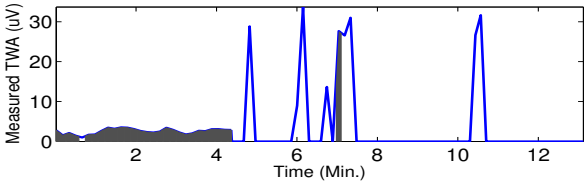
3.5.7 Summary

In summary, we applied the proposed Adaptive TF quantification to TWA quantification, and called the technique as Adaptive SM. In Adaptive SM, first we transformed the aligned T waves from the beat domain to a TF plane using Adaptive TFD which provided a high time and frequency resolution of TWA variations. Next, the Adaptive TFD constructed above was quantified at the 0.5 cpb energy to estimate the TWA. The proposed Adaptive SM was evaluated in comparison with two commonly used approaches, i.e. SM and MMA, under a wide range of data non-stationary conditions and noise. In the presence of data non-stationarity, such as, phase reversal and ectopic, the Adaptive SM significantly performed more accurately compared to the classic methods. Both the Adaptive SM and MMA techniques successfully tracked the changing TWA magnitude, while SM did not correctly followed the changes. All the three techniques showed a similar behavior under changing heart rate with a slight change in the measured TWA.

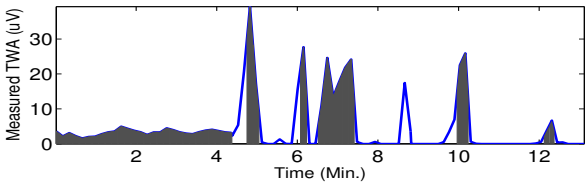
In the presence of periodic, Gaussian and physiological noise, Adaptive SM was more robust in discriminating simulated TWA from noise compared to SM and MMA. MMA tended to falsely



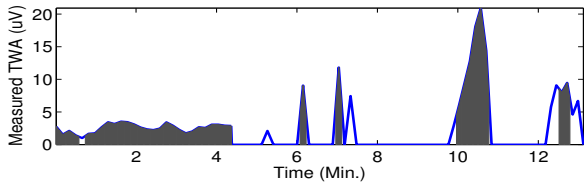
(a)



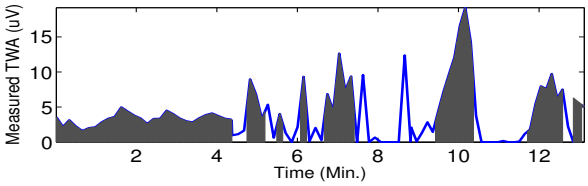
(b)



(c)



(d)



(e)

Figure 3.21: ECG recording (lead V2) in one patient during atrial pacing where frequent ectopy develops during pacing rates of 110 and 120 bpm. Adaptive SM and SM are compared under these conditions. (a) Heart rate during atrial pacing. (b) TWA measurement using SM. (c) TWA measurement using Adaptive SM. (d) TWA measurement using SM after manual replacement of the ectopic beats. (e) TWA measurement using Adaptive SM after manual replacement of the ectopic beats. The shaded area illustrates significant TWA signal above ambient noise ($K_{score} > 3$).

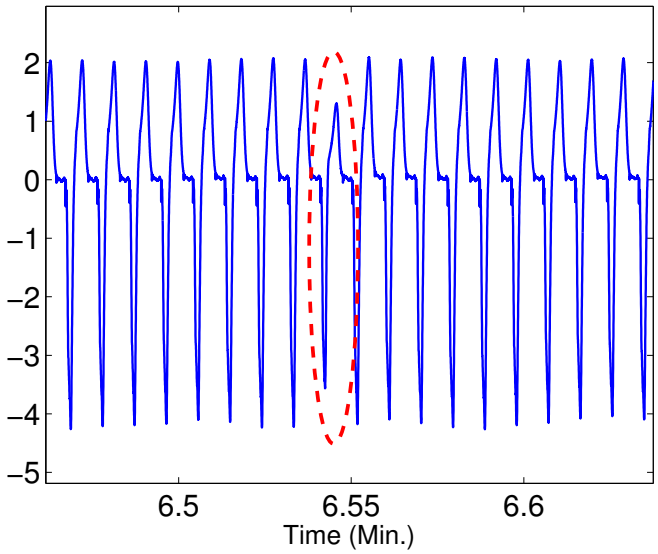


Figure 3.22: In the pre-processing stage, all the ectopic beats with a QRS morphology template correlation less than 0.85 are replaced with an average beat. However, in practice, some ectopic beats will not be detected. As shown in this figure, the ectopic beat at 6:54 min. is not automatically detected by our algorithm as its correlation with the average beat is higher than our pre specified threshold.

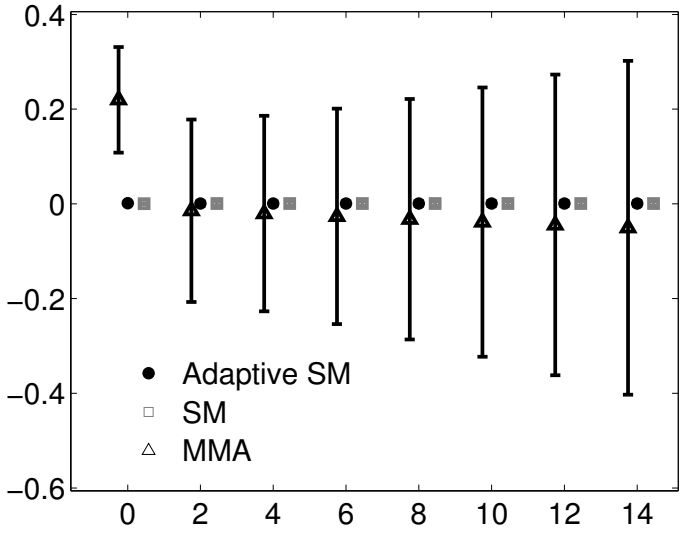


Figure 3.23: TWA measurement error (mean±SD) in ambulatory ECGs using SM, Adaptive SM and MMA as a function of TWA magnitude.

detect or overestimate simulated TWA in synthetic ECGs and ambulatory ECG recordings, while SM underestimated the TWA in noisy recordings. In all experiments, Adaptive SM was able to accurately estimate the TWA compared to SM and MMA.

Adaptive SM was compared to SM in an experiment with invasive ECG recordings as a real case of TWA. The purpose of this part of the study was to show the effect of unreplaced ectopic beats on alternans detection. We knew that the patient had a high TWA, but there was no quantified value. The results showed that Adaptive SM successfully measured the TWA before and after manual removal of ectopic beats, while SM missed the TWA unless we manually replaced the ectopic beats.

Experiments performed with synthetic and real ECGs demonstrated the potential of the Adaptive SM in important clinical implications for improving the accuracy of TWA measurement in ambulatory ECG recordings, particularly when residual noise remains after ECG preprocessing that may confound signal detection. Table 3.1 summarizes our evaluation performed in this chapter. The more is the number of stars at each property means that the method is more desirable.

Table 3.1: Desirable Properties for TWA Quantification. The more are the number of the stars at each property indicates that the method is more desirable with respect to that specific property.

Property	Random Noise	Periodic Noise	Heart-rate Change	TWA Change	Ectopics	Phase Reversal
MMA	*	*	**	**	**	*
SM	**	**	***	*	*	**
Adaptive SM	***	***	**	***	***	***

It can be seen that MMA provided more accurate signal quantification in the presence of data non-stationarity; however it was not robust in the presence of noise contamination. On the other hand, the SM was more robust in noisy conditions, but it was less accurate in the presence of TWA non-stationarity. However, it can be seen that Adaptive SM is superior to SM and MMA in terms of tracking the non-stationarity of the T waves and preserving its robustness in the presence of noise. Therefore, this technique has a high potential of technology transfer to replace the current invasive

testing to identify patients at high risk of SCD. Additionally, it may lead to the development of novel implications in cardiac monitoring that can benefit global health care.

3.6 Chapter Summary

This chapter presented a quantification technique to obtain features which successfully represent a known pattern. Adaptive TF signal representation was used as a desirable tool to represent both the long-term and non-stationary characteristics of a signal. Then, we adopted a suitable TF feature extraction technique to quantify the known characteristics of signals. The proposed Adaptive TF analysis was employed to provide a flexible representation of amplitude varying structures with an excellent TF resolution without cross-terms. This representation was successfully used to quantify TWA as a risk indicator of SCD. The proposed method accomplished to compensate the limitations of current TWA quantification methods and achieve a highly robust and non-stationary adaptive technique.

In this chapter, we focused on the quantification and detection of known patterns that belonged to the signal's nature. However, there are scenarios where the structures of interest are embedded into the signal. In these applications, the TF features are deliberately inserted into a signal, and our goal is to detect these known and embedded signatures. The next chapter, we further investigate such implications of TF feature detection.

Chapter 4

EMBEDDED TF FEATURE DETECTION

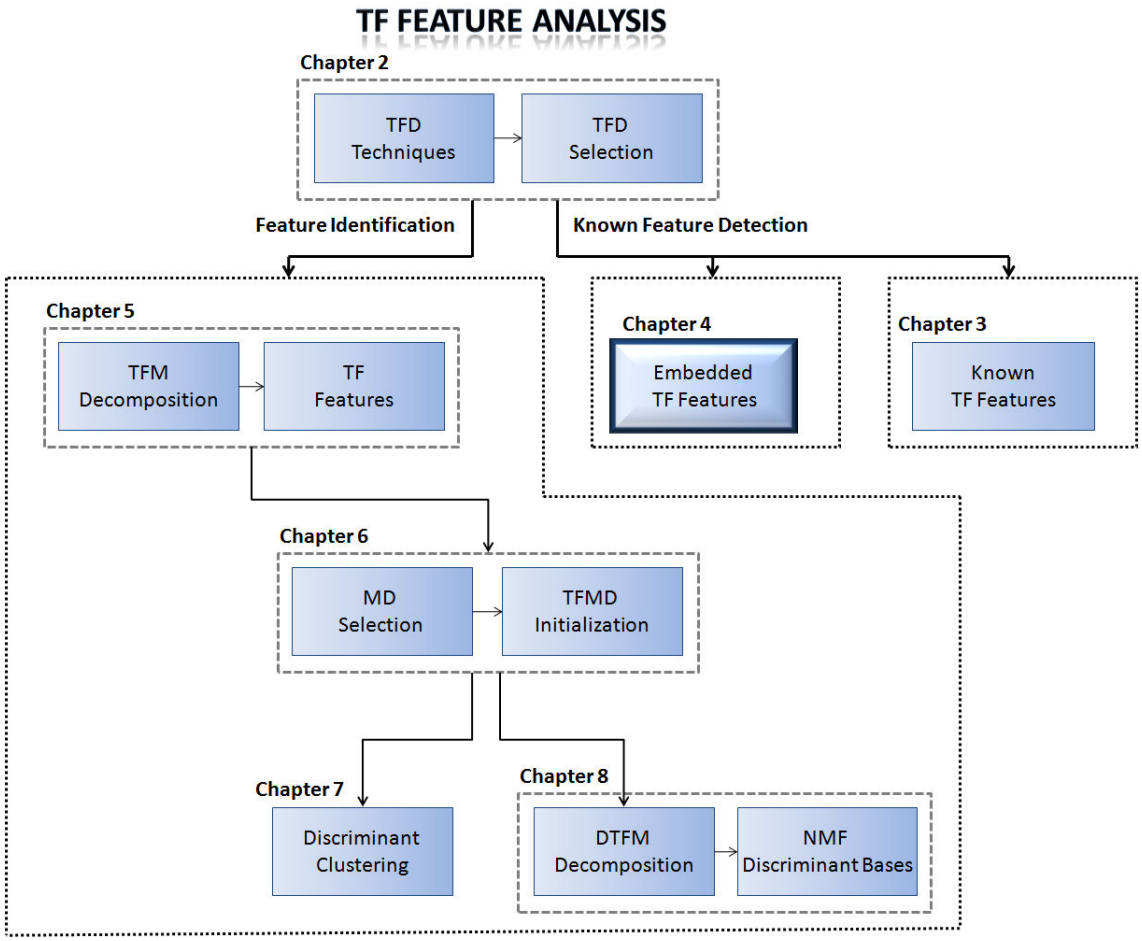
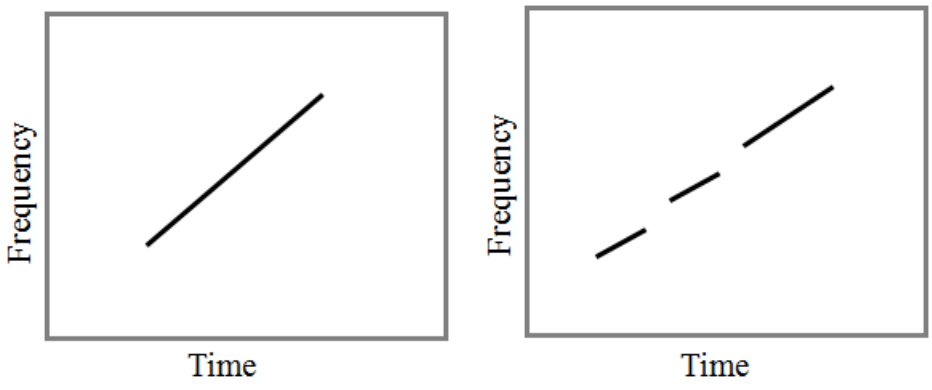


Figure 4.1: Chapter 4 - Embedded TF quantification and detection.



(a) (b)

Figure 4.2: (a) A linear chirp in TF plane. (b) The corrupted chirp in TF domain.

4.1 Motivation

THE challenges we faced so far aimed to detect the known discriminating patterns. In the previous chapter, these patterns were known and belonging to the signal's structure, while the present chapter focuses on the quantification of known TF features that are deliberately embedded into a signal. The patterns of interest can be embedded as known structures in the spectral characteristics of a given signal. An example of such known spectral characteristics is a linear frequency variation of a chirp signal. Fig. 4.2(a) illustrates the structure of a chirp in a TF plane. The unique TF structure of this signal makes the pattern detection process very simple. It is enough to extract the slope of the chirp, or its initial and final frequencies, so we can easily detect the presence of a chirp pattern. However, in most applications, the detection task is not as easy as it might sound since the known structure is usually corrupted by the presence of noise or other outliers in the signal. Fig. 4.2(a) shows a perfect chirp, while Fig. 4.2(b) displays the same chirp with some corruptions. It can be seen that due to the noise in this signal, some parts of the chirp are missing. In order to successfully detect this chirp, we require to develop a pattern quantification approach that extracts robust features that represent the TF varying structure of the chirp even under the presence of noise corruptions.

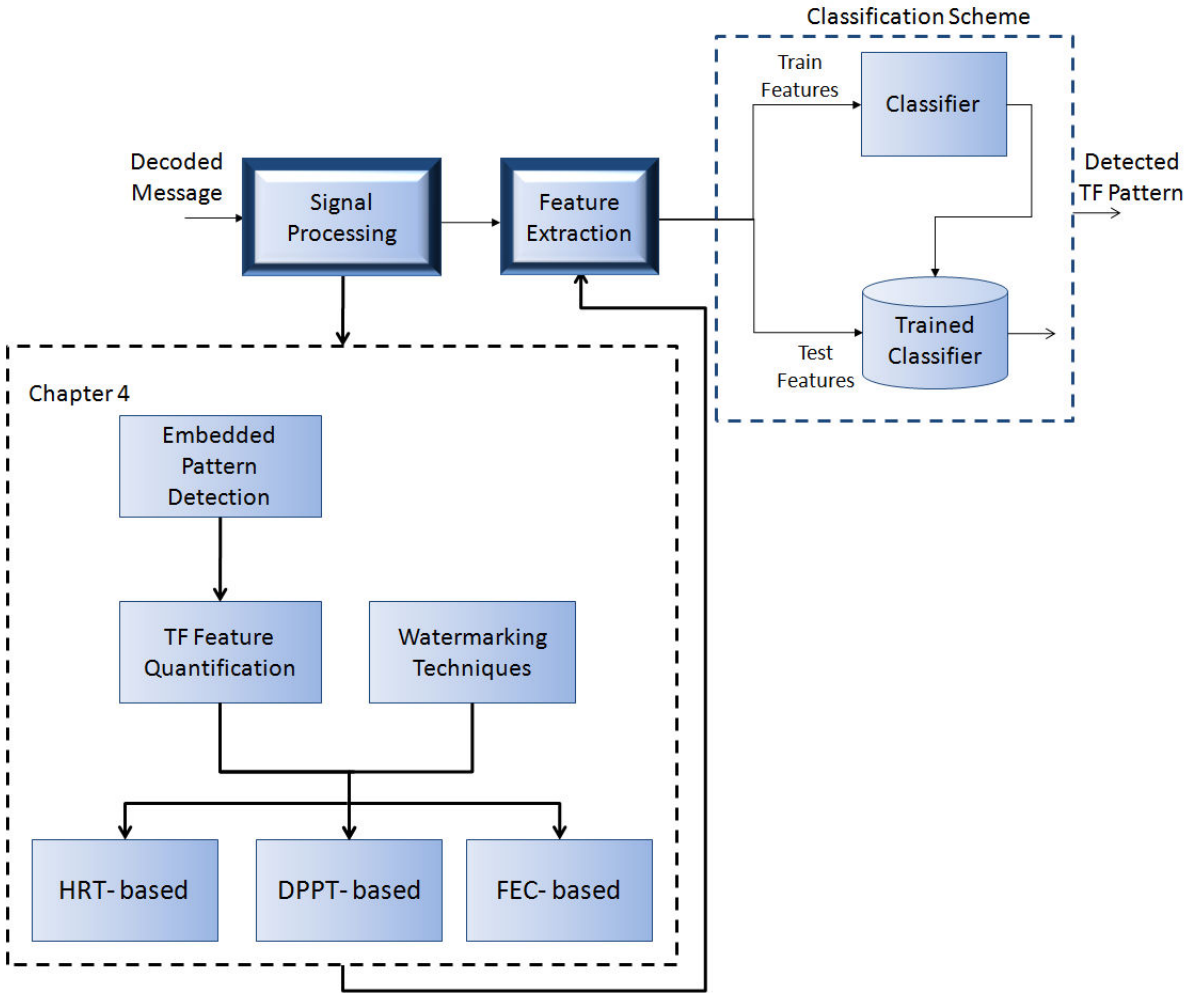


Figure 4.3: Chapter 4 - Embedded TF Feature detection

Linear chirps are signals with time-varying frequency, which are present in many areas of science and engineering. If the right TF quantification technique is utilized, the unique structures of chirps allow us to successfully detect them even in the presence of a very high bit-error-rate (BER). Having said this, the present chapter aims on quantification of known TF features with main focus on chirp detection applications. Fig. 4.3 shows the schematic of this chapter's contribution. First, we explain the characteristics of chirp signals. Next, the techniques to quantify the known TF structure of chirps are introduced. Our desirable TF features are robust to noisy conditions; i.e., the obtained TF features remain unchanged if the chirp signal is noisy. Finally, in order to

evaluate the proposed TF features, we apply them to multimedia security applications, and study the obtained TF features under severe noisy conditions.

4.2 Embedded TF Signatures

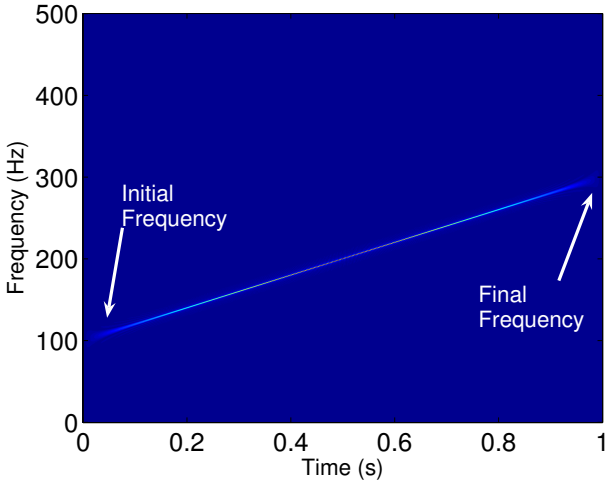
Linear chirps with different slopes can be interpreted as signals with different messages. Fig. 4.4 displays two of such messages in the TF plane. As it can be seen in this figure, each chirp is displayed in the TF domain as a line with different slopes. The chirp in Fig. 4.4(a) starts from frequency of 100 Hz, and ends with the frequency of 400 Hz, while the other chirp, shown in Fig. 4.4(b) starts with 100 Hz but ends at 300 Hz.

In the above example, if we develop an appropriate feature extraction tool to represent the slope of each chirp signal, the extracted features will be enough to detect the hidden message. However, due to signal manipulations, the chirp signal might not be a continuous line in the TF plane. Fig. 4.5 shows a chirp-based message in TF domain. This message has been corrupted with 20% bit-error-rate (BER) [1]. An efficient TF quantification tool has to quantify the characteristics of the noisy chirp in this figure in a way that the TF features robustly represent the characteristics of the chirp. Next section explains the techniques that can be used to quantify a chirp's characteristics while ignoring the effect of noise on the chirp. Having said this, in the rest of this chapter we develop such an accurate and robust TF feature quantification technique.

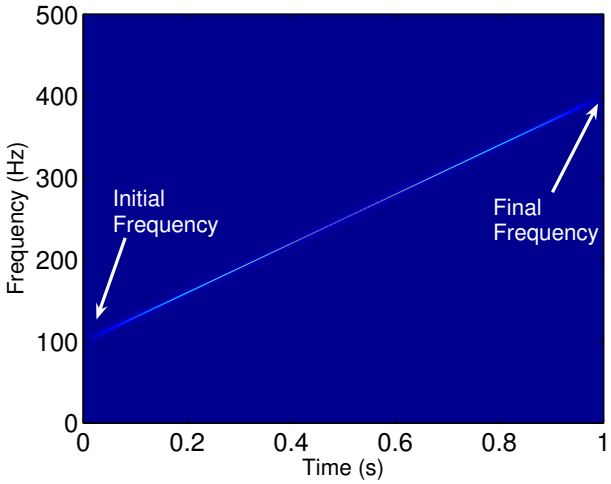
4.3 TF Feature Quantification Techniques

Knowing the fact that the TF structure of our interest is linearly frequency modulated, this section explains two TF feature quantification approaches: Hough-Radon Transform (HRT), and Discrete polynomial phase transform (DPPT). Both approaches attempt to detect the chirp parameters, where HRT detects the slope of the chirp in TF domain, while DPPT is a temporal technique that estimates the varying phase of the chirp over time.

[1] Bit-error-rate (BER) calculates the percentage of the bits that have been corrupted over the entire message.



(a)



(b)

Figure 4.4: Two chirp signals with different TF characteristics are shown in TF plane. (a) Start and ending frequency: 100 Hz and 400 Hz, respectively. (b) Start and ending frequency: 100 Hz and 300 Hz, respectively.

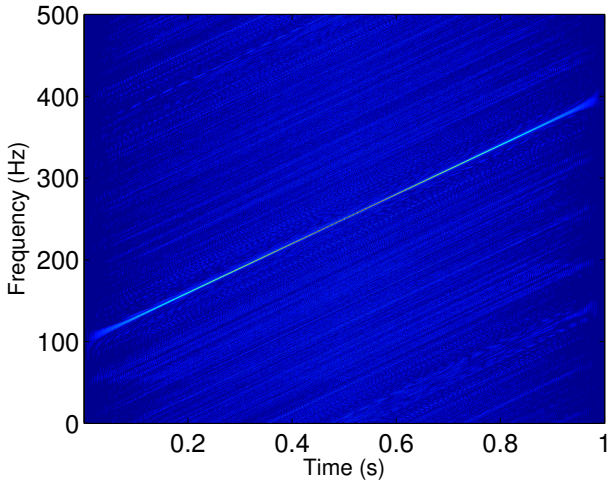


Figure 4.5: A chirp signal with BER of 20%.

4.3.1 Hough-Radon Transform (HRT)-based TF Features

Hough-Radon Transform (HRT)

The HRT is developed by Rangayyan and Krishnan [50] to detect linear and non-linear frequency modulated signals from the image of the TFD. The HRT is a combination of the Hough transform (HT) and Radon transform (RT). Line detection by the HT is performed by quantizing the parameter space, and incrementing the accumulator cells by one value for each pixel on a straight line. Although the HT is a robust method which is insensitive to missing data, the major drawback with the HT is that it is defined for a binary image and cannot be applied to a gray-level image. On the other hand, RT is a line detector which is applicable in gray level images. RT identifies the straight lines in an image by adding up the pixel values in the given image along a straight line, but the space quantization of HT is not possible in RT. Given the advantages and the drawbacks of the RT and the HT individually, it is appropriate to combine the RT with the HT to identify complex TF features with varying gray level or intensity. In HRT, instead of incrementing each accumulator cell by one value for each pixel on the straight line, the energy (or gray scale value) of each pixel is added to the accumulator.

Let us consider the problem of detecting a chirp represented by a TFD $W(n, \omega)$. The TFD

is treated as a gray scale image and the chirp is identified by detecting a straight line in the TFD image. The straight line is represented by

$$x\cos\theta + y\sin\theta = \rho. \tag{4.1}$$

where θ, is the angle of the ray path of integration, ρ, is the distance of the ray path from the center of the image. In the implementation of the HRT, the parameter space (ρ, θ), also known as the HRT space, is bounded in $\theta \in [0, \pi]$ and ρ, by the greater of rows and columns (say rows) $\pm rows/\sqrt{2}$. The HRT space is divided into accumulator cells and the cell at coordinates (i, j) with accumulator value $A(i, j)$ corresponds to the partition of the space associated with the parameter coordinates (θ_i, ρ_j). In other words, the transform value at some point (ρ_0, θ_0) in the Hough space contains the total energy in the pixels that satisfy the parameter constraint equation. Therefore, the cell with the highest value determines the parameters of the line.

HRT Features

The HRT is a tool to detect the pixels that form a parametric constraint of either a line or curve in a gray level image [50]. If the chirp signal is denoted with m , the TFD of the signal is constructed as follows:

$$\mathbf{I}_{(M\times N)} = \text{TFD}\,(m) \tag{4.2}$$

In order to achieve a good detection performance, Wigner-ville distribution (WVD) is used to represent the TFD of the signal with a high TF resolution. Then, the HRT is applied on the TF image **I**:

$$\text{HRT}(\theta) = \sum_{t=1}^{N}\sum_{f=1}^{M} \mathbf{I}(f,t)\delta\left(\rho - (t\text{cos}\theta + f\text{sin}\theta)\right), \tag{4.3}$$

$$\theta \in [0,\pi]\,, \rho \in [-K, K]\,, \tag{4.4}$$

where, HRT(θ) is the HRT of the TF image, and represents the summation of the image along the line associated to ρ and θ in the HRT space, and K is the number of bins in HRT space.

The HRT space is quantized into $K \times K$ cells, and the cell with the highest accumulator value, (k_0, k_1), corresponds to the parameters of the HRT constraints which are the initial and final frequencies of the corrupted chirp m:

$$\hat{f}_i = (k_0 - 1) f_s / (2N), \tag{4.5}$$

$$\hat{f}_e = (k_1 - 1) f_s / (2N), \tag{4.6}$$

where, f_s is the sampling frequency of the chirp signal, N is the signal length, and $\hat{f}_i$ and $\hat{f}_e$ are the initial and final frequency estimations. Fig. 4.7 shows the WVD and the HRT space of a linear chirp received at a BER of 21%. The global maximum in HRT space indicates a correct estimation of the initial and final frequencies of the linear chirp, and from the recovered chirp in Fig. 4.7 (a), it is evident that even in a very high BER condition, the HRT is able to successfully extract the TF characteristics of a chirp signal.

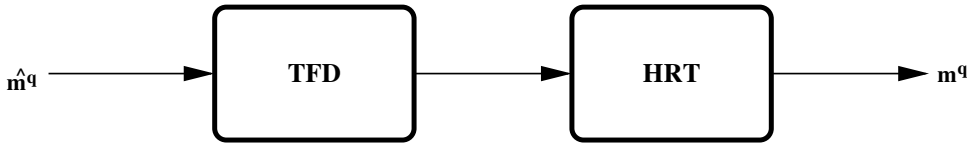


Figure 4.6: HRT-based TF feature extraction of a signal with linear time-varying frequency.

4.3.2 Discrete polynomial phase transform (DPPT)-based TF Features

The DPPT has been extensively studied in recent years [51, 52, 53]. It is a parametric signal analysis approach for estimating the phase parameters of constant-amplitude polynomial phase signals. The DPPT operates directly on the signal in time domain and is a computationally efficient method. The principle of DPPT is discussed as follows:

Discrete polynomial phase transform (DPPT)

The discrete version of a polynomial phase signal can be expressed as:

$$x(n) = b_0 \exp\left\{ j \sum_{q=0}^{Q} a_q (n\Delta)^q \right\} \tag{4.7}$$

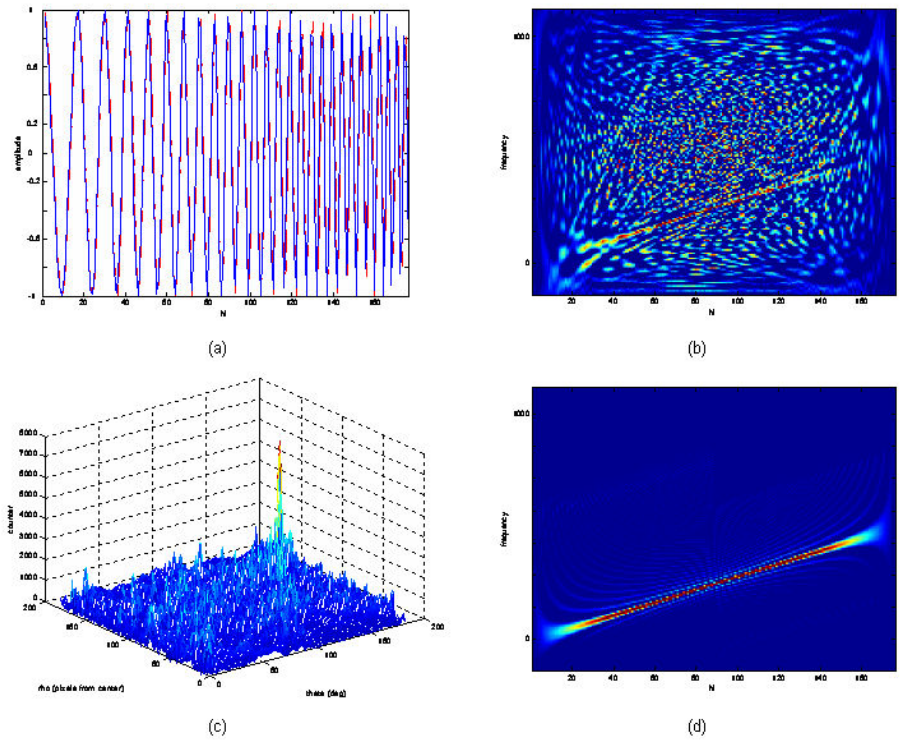


Figure 4.7: HRT performance at BER of 21%. (a) The reference and extracted chirps in time domain. (b) The WVD of the detected chirp with 21% BER. (c) The HRT of the detected chirp in Hough space; and (d) the WVD of the chirp using parameters extracted by HRT.

where Q is the polynomial order ($Q = 2$ for chirp signal), $0 \leq n \leq N-1$, N is the signal length and Δ is the sampling interval.

The DPPT algorithm introduces operators, which applying the Q-order operator to a constant-amplitude polynomial-phase signal of the same order transforms the broadband signal into a single tone signal with frequency related to a_q. By using a Fast Fourier transform (FFT) calculation, the position of this spectral line at frequency ω_0, which provides an estimate of the coefficient $\hat{a}_q$, is determined. After $\hat{a}_q$ is calculated, the order of the polynomial is reduced from q to $q-1$ by multiplying the signal with $\exp\{-j\hat{a}_q(n\Delta)^q\}$. The next coefficient $\hat{a}_{q-1}$ is estimated the same way by taking DPPT of the polynomial phase signal of order $q-1$. This procedure is repeated until all the coefficients of the polynomial phase are estimated from the highest order to the lowest order phase coefficient. Since in this study we utilize linear chirps, in the rest of this material, we focus

only on the DPPT of order $M = 2$. DPPT operators are defined as [51]:

$$\mathcal{DP}_1 [x(n), \tau] := x(n) \tag{4.8}$$

$$\mathcal{DP}_2 [x(n), \tau] := x(n)x^*(n - \tau). \tag{4.9}$$

where $x(n)$ is a fixed amplitude linear phase signal and τ is a positive number. The coefficients a_2 and a_1 are estimated by applying the following formula:

$$\hat{a}_2 = \frac{1}{2(\tau_2 \Delta)} argmax_\omega \left\{ |\mathrm{DPPT}_2 [\mathrm{x(n)}, \omega, \tau]| \right\}, \tag{4.10}$$

$$\hat{a}_1 = argmax_\omega \left\{ |\mathrm{DPPT}_1 [\mathrm{x(n)}, \omega, \tau]| \right\}, \tag{4.11}$$

$$\hat{a_0} = phase \left\{ \sum_{n=0}^{N-1} x(n) exp \left\{ -j \left(\hat{a}_1 (n\Delta) + \hat{a}_2 (n\Delta)^2 \right) \right\} \right\} \tag{4.12}$$

where

$$\mathrm{DPPT}_1 [\mathrm{x(n)}, \omega, \tau] = \mathcal{F} \left\{ \mathcal{DP}_1 [\mathrm{x(n)}, \tau] \right\}, \tag{4.13}$$

$$\mathrm{DPPT}_2 [\mathrm{x(n)}, \omega, \tau] = \mathcal{F} \left\{ \mathcal{DP}_2 [\mathrm{x(n)}, \tau] \right\}, \tag{4.14}$$

The estimated coefficients are used to synthesize the polynomial phase signal:

$$\hat{x}(n) = exp \left\{ j(\hat{a}_0 + \hat{a}_1 n + \hat{a}_2 n^2) \right\} \tag{4.15}$$

DPPT Features

Fig. 4.8 shows the TF quantification using DPPT method. A chirp signal is defined as shown in

Figure 4.8: DPPT-based TF feature extraction of a signal with linear time-varying frequency.

the following equation:

$$\begin{aligned} \hat{m}(n) &= \sin \left(\pi (f_{1m} - f_{0m}) \tfrac{n^2}{N} + 2\pi f_{0m} n \right) \\ &= \sin \left(a_2 n^2 + a_1 n + a_0 \right), \end{aligned} \quad n = 1, ..., N \tag{4.16}$$

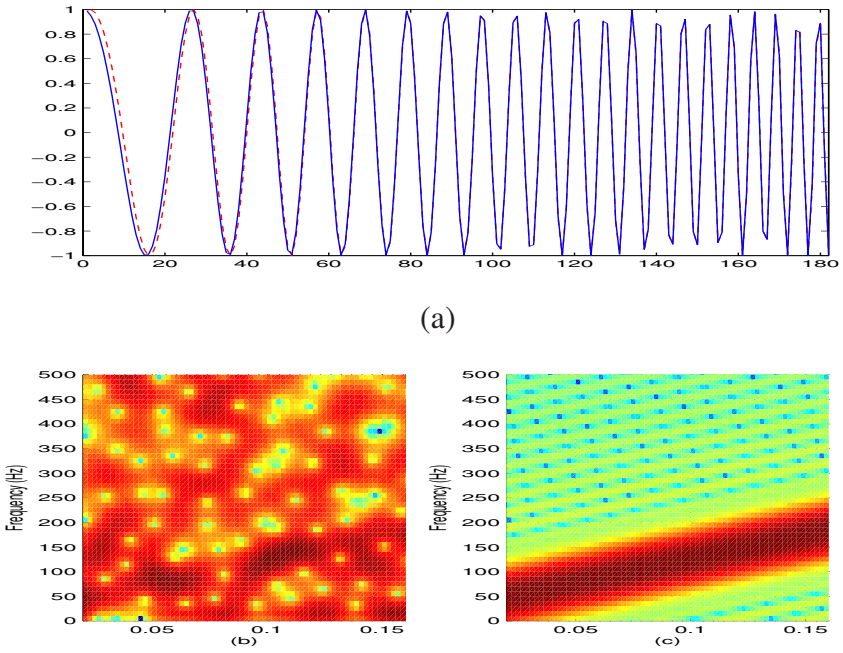


Figure 4.9: (a) The original and noisy chirps at BER of 21% while the solid line is the original message and the dashed one is the reconstructed chirp. (b) Spectogram of the noisy chirp; and (c) shows the spectogram of the estimated chirp reconstructed using the DPPT-based features.

Applying DPPT technique on this signal, we achieve an estimate of signal's phase coefficients $\hat{a}_2$, $\hat{a}_1$ and $\hat{a}_0$. Fig. 4.9 shows the reference and the extracted chirp under BER of 21%. Although there are discontinuities in the time frequency domain of the chirp signal, DPPT completely recovers the message, and successfuly extracts robust TF signatures.

The fact that DPPT technique is applied on the time domain signal makes the DPPT method a faster quantification technique compared to the HRT-based method, which requires to transform the signal into a TF plane.

4.4 Experiment: Multimedia Security

In order to verify the accuracy of the mentioned feature extraction techniques above, in this section we apply the techniques to detect known TF features embedded in multimedia security application.

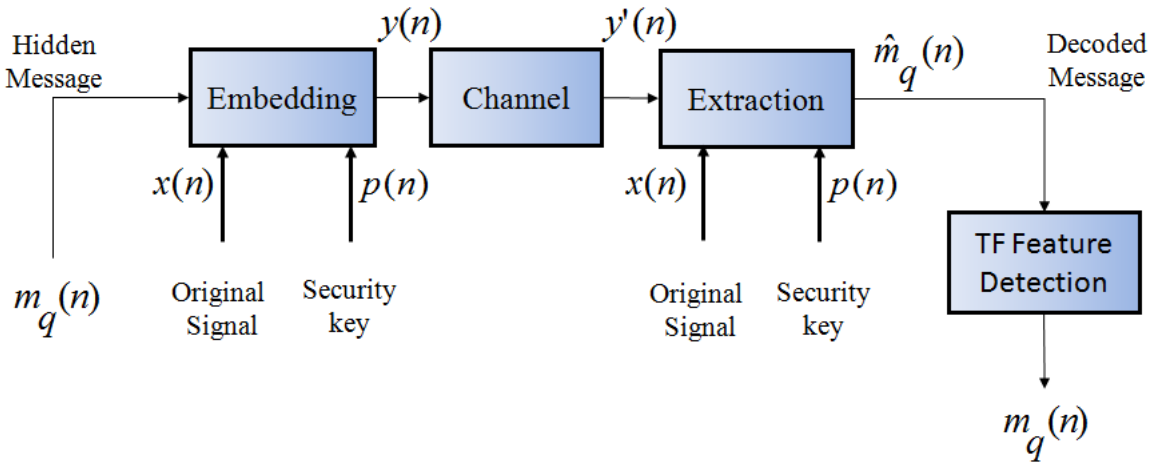


Figure 4.10: Block diagram of a message hiding system.

4.4.1 Background

In recent years, the digital format has become the standard for the representation of multimedia content. Today's technology allows the copying and redistribution of multimedia content over the Internet at a very low or no cost. This has become a serious threat for multimedia content owners. Therefore, there is significant interest to protect copyright ownership of multimedia content (audio, image, and video). Watermarking is the process of embedding external data into the host signal for identifying the copyright ownership. The embedded data characterizes the owner of the data and should be extracted to prove ownership. Besides copyright protection, watermarking may be used for data monitoring, fingerprinting, and observing content manipulations.

Fig. 4.10 displays the schematic of a message embedding system. As it can be seen in this block diagram, the system consists of three main stages: embedding, extraction and detection. In embedding, a message, $m_q(n)$, is embedded in the original signal, $x(n)$, to produce a coded content, $y_q(n)$. The main application of a message embedding system is in multimedia security purposes in which the hidden message may contain information about the owner of the content and the access conditions of the multimedia content, and is combined with a key code $p(n)$ to preserve the security of the data to those only who have access to the key code. The coded content is passed through a block called channel. The channel is referred to the possible signal processing

manipulations that might be applied to the chirp message by users, such as data compression techniques or filtering process. The channel also includes the illegal attempts of the users to remove the hidden message. The extraction stage decodes the embedding process using the security key. However, due to the signal processing on the coded content, even using the most robust embedding techniques, there will be some bit errors in the decoded message, $\hat{m}_q(n)$. The final stage in Fig. 4.10 attempts to detect the embedded message from the degraded message. This stage uses the knowledge about the structure of the embedded message to reconstruct the original message.

Various digital watermarking methods have been researched by many authors in the past years. The watermarking techniques differ depending on their applications and characteristics such as invisibility, robustness, security and media category. These methods are classified either by the processing domain or the type of the watermark message [54]. The processing domain, where the insertion and extraction of watermark takes place, is usually spatial domain [55] or frequency domain [56], and few of them are done in the joint time-frequency domain [57]. The watermark message can be either a random signal, i.e. pseudo noise sequence, a Gaussian random sequence or an image such as the ones in the form of binary image, stamp, and logo [58, 59]. Enhancement of either the embedding and extraction techniques or the selection of the right watermark message can develop a successful watermarking algorithm that satisfies the desirable criteria mentioned earlier in this chapter.

Watermarking schemes can be classified into blind and non-blind schemes according to the watermark detection procedure used at the receiver [60]. The difference between these two schemes is that non-blind schemes detect the watermark using the original signal whereas blind schemes do not use the original signal for watermark detection. Although non-blind schemes are more robust in detecting watermarks, the multimedia industry appears to prefer the blind schemes due to their practicality.

4.4.2 Embedding Techniques

The watermarking literature also describes a second classification made according to the amount of data that a watermark carries. In particular, there are two classes considered. The first class in-

cludes the one-bit watermarking techniques [61], which detect only the presence of the watermark. The second class of techniques not only detects but also extracts the embedded watermark message. In [62], the proposed watermark embedding algorithm projects the audio signal's frequency subbands onto a secret key. In [63], the authors present an algorithm which embeds watermarks in time-domain using the energy of consecutive audio blocks. An algorithm embedding the watermark as noise into Fourier coefficients is proposed in [64]. Gang et. al. propose an algorithm considering the effects of MP3 compression [65]. All these algorithms embed and extract multiple watermark bits. As a result of signal manipulations, some message bits extracted by the detector may be in error, potentially resulting in the detection of the wrong watermark message.

There is a trade-off between the imperceptibility of the hidden signature and the robustness of the feature detection process. If we increase energy of the embedding stage, the message will be robust to any signal manipulations, but it will also damage the quality of the base content; and vice versa. A successful message embedding system should satisfy the following requirements [66]:

- Unobstructive – that is perceptually imperceptible, when embedded into the host signal.
- Discreet – undetectable to prevent unauthorized removal.
- Robust – the embedded message should remain intact in the host signal when subjected to intentional removal attacks and common signal processing manipulations.
- Easily extractable – authorized users must easily detect the hidden signature.

During the last decade, many watermarking methods have been proposed in literature which all of them were common in pursuing the robustness of watermarking; amongst them, quantize-and-replace strategy [67][68], additive spread-spectrum-based method [60] and quantization index modulation (QIM) [69] are the most popular information-embedding algorithms. The main goal of this study is to evaluate the performance of the watermarking system using TF signatures and different chirp detection tools. Therefore, any reliable watermarking technique fulfills our objective. In the entire experiments, we utilize spread spectrum technique as the watermarking method which is a robust and well-known technique [70].

4.4.3 Message Selection

Beside the attempts to design a robust embedding and extracting method, some papers in literature have focused on applying a watermark detection technique on the decoded watermark signature, so that correction of the bit error rate (BER) after extracting the watermark will be possible. One example of such approaches is encoding the watermark signature with a forward error correction (FEC) scheme before embedding the watermark to the document [71]. In communication technologies, FEC schemes have been used to protect the data from channel noise. Before transmission of the data, some redundant bits are inserted into the signal in a way that these additional bits can be useful in the receiver for detecting and correcting the bit errors occurred during transmission. Similarly, in watermarking and fingerprinting applications, before embedding the message, it is encoded using a FEC encoding technique. After extraction of the embedded message, the FEC decoder is applied to compensate the bit error, and successfully detect the embedded watermark.

To enhance both the robustness and invisibility of the embedded message, our approach is to embed a known-structured message into the multimedia content. Once the message is extracted from the multimedia, we apply a technique to detect the hidden structure from the decoded message. In this algorithm, the detection process behaves as a TF feature detection that enhances the message recovery even if the decoded message contains some bit errors.

In most the applications, we do not have the luxury of choosing the structure of the signal to be detected. However, in known embedded feature extraction, we can define what should the structure of the interest be like. One of the works in this direction is shaping the embedded messages into linearly frequency modulated signals called chirps. This idea is suggested in [72, 73] where linear chirps are embedded as the watermark messages instead of randomly generated signals. The following properties of chirps provide us a strong motivation to focus on chirp-based TF feature detection:

- Chirps are time varying frequency signals and they can be optimally detected in a TF plane.
- Different chirp rates, i.e., slopes on the TF plane, could represent different messages such that each slope corresponds to a different message.

- The technique can reliably extract hidden messages even in the presence of bit errors caused by signal or channel manipulations.
- Provides an effective signature for checking the presence of a structure in a given signal.

4.4.4 Spread Spectrum

In this study we use spread-spectrum watermarking scheme which is a correlation method that embeds pseudo-random sequence and detects watermark by calculating correlation between pseudo-random noise sequence and watermarked signal. Spread-spectrum scheme is the most popular scheme and has been studied well in literature [70]. The spread-spectrum method can be applied in time domain or transformed frequency domain. We utilized discrete cosine transform (DCT) coefficients, which are widely used in compression applications and are easier to impose human visual system (HVS) constraints on them. Fig. 4.11 shows the block diagram of the watermarking embedding and extraction schemes [73].

Watermark Embedding

First, the DCT coefficients of each 8 X 8 block of original image is calculated; then to find the regions of the image that the DCT coefficients can be changed without producing large visual artifacts, the *just noticeable difference* (JND) [74] of each block is computed. JND paradigm finds the perceptual entropy of the image based on HVS, and determines the perceptually significant regions to embed the watermark. Matrix $X_{u,v,b}$ represents the DCT coefficients of block b of the image. Then, the watermark message is spread with a shifted version of a random PN sequence:

$$\mathbf{w} = \sum_{k=0}^{N-1} m_k^q \mathbf{p_k}, \tag{4.17}$$

where m_k^q is bit k of the watermark message $\mathbf{m^q}$ of length N, and $\mathbf{p_k}$ is the cyclic shifted version of the PN sequence. The PN sequence is the key of the watermarking and preserves the security of the watermarking method; since illegal extractors do not have access to the watermarking key, they are not able to extract the hidden watermark. In addition, the spread message satisfies the robustness of the watermarking and makes the embedded message less imperceptible. Next, the

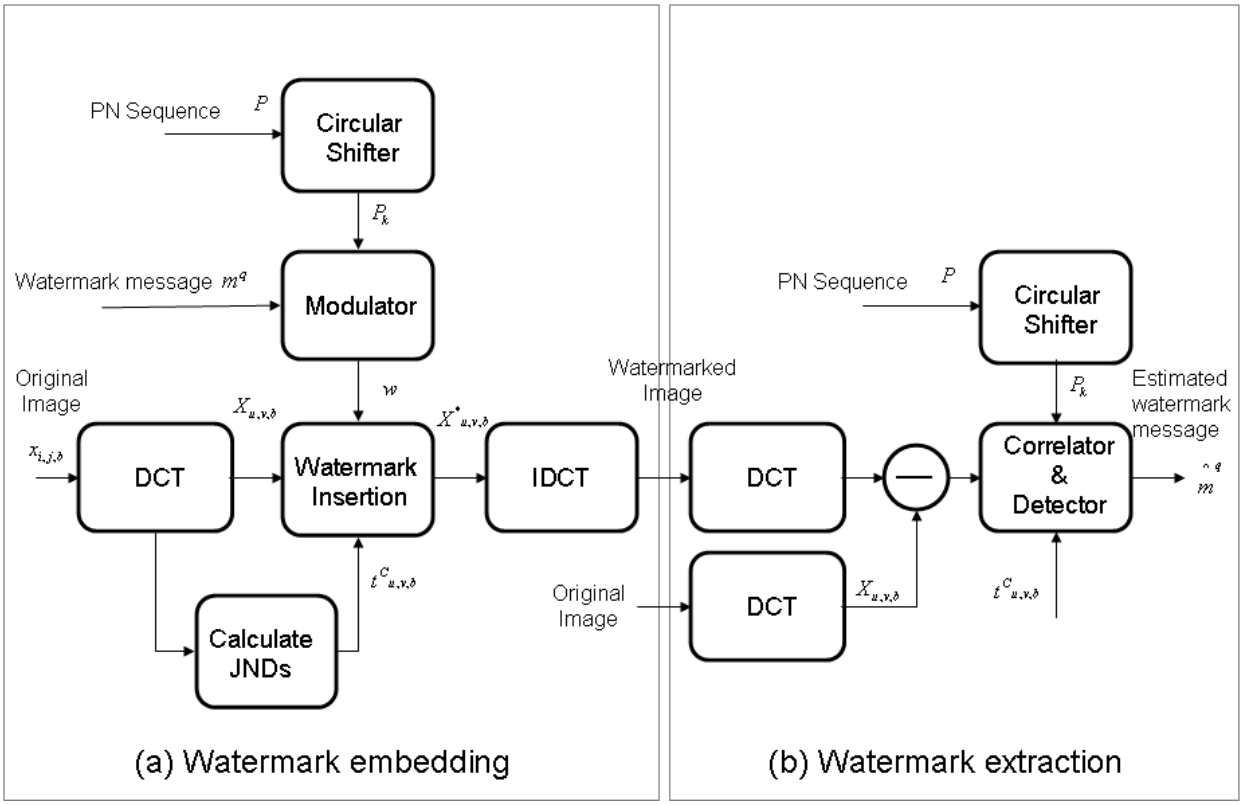


Figure 4.11: Detection and extraction block diagram of the watermarking method.

spread watermark message is embedded to the DCT domain of the original image $X_{u,v,b}$ in regions that the DCT coefficients are higher than the computed JND $t^C_{u,v,b}$:

$$X^*_{u,v,b} = \begin{cases} X_{u,v,b} + t^C_{u,v,b} w_{u,v,b}, & \text{if } X_{u,v,b} > t^C_{u,v,b}; \\ X_{u,v,b}, & \text{otherwise} \end{cases} \tag{4.18}$$

Finally, the watermarked DCT coefficients $X^*_{u,v,b}$ are converted back to time domain by IDCT to get the watermarked image.

Watermark Extraction

In watermark extraction scheme the difference between the watermarked and original DCT coefficients is computed. Then, using the watermark key the received wideband noise vector $\hat{\mathbf{w}}$ is despread in the regions that the original image was marked in the embedding process:

$$\hat{w}_{u,v,b} = \frac{X_{u,v,b} - \hat{X}^*_{u,v,b}}{t^C_{u,v,b}} \tag{4.19}$$

$$\hat{\mathbf{w}} = \begin{cases} \hat{w}_{u,v,b}, & \text{if } X_{u,v,b} > t^C_{u,v,b}; \\ 0 & \text{otherwise} \end{cases} \tag{4.20}$$

The sign of the expected value of $\langle \hat{\mathbf{w}}, \mathbf{p}_k \rangle$ depends only on the embedded watermark bit m^q_k. Hence using a detection rule, the watermark bits can be estimated.

$$\hat{m}^q_k = \begin{cases} +1, & \text{if } \langle \hat{\mathbf{w}}, \mathbf{p}_k \rangle > 0; \\ -1, & \text{if } \langle \hat{\mathbf{w}}, \mathbf{p}_k \rangle < 0. \end{cases} \tag{4.21}$$

Repeating the estimation process for all the watermark bits, the watermark message is extracted. As mentioned earlier, because of the intentional and non-intentional signal processing, the received message will suffer with some BERs. In the next section, taking the advantages of the known structure for the watermark message, we try to compensate and correct these bit errors.

4.4.5 Watermark Quantification Techniques

The received watermark message bits $\hat{m}$ is detected using the watermark detection scheme explained in Section 4.4.4. Since the embedded watermark is a linearly frequency chirp with initial and final frequencies of f_i and f_e respectively, the received message can be represented as:

$$\begin{aligned} &\hat{m}(n) = \sin\left(\pi(f_i - f_e)\tfrac{n^2}{N} + 2\pi f_i n\right) \\ &n = 1, ..., N \end{aligned} \tag{4.22}$$

Because of signal manipulations of the watermarked image, or illegal attempts of the users to remove the watermark, there may have been some errors in the received bits. To compensate these errors, and extract the embedded message correctly, we take the advantage of the known chirp structure in the watermark message. There are various methods available for detection of chirps in the time domain, joint TF domain and the ambiguity domain [75, 76, 77]. The common techniques for linear chirp detection include the HRT [50], and the DPPT [51, 53] as explained in the previous section. HRT and DPPT are two chirp quantification methods that have the capability to extract robust features that are unchanged even under presence of bit error in the received chirp. The HRT or DPPT techniques are applied to the decoded message as shown in Figs. 4.6 and 4.8. Once the TF features are calculated, we compare them with the characteristics of the possible messages, and select the message with closest structure as the hidden message.

HRT-based Method

Depending on the number of possible cells in the HR domain, a certain number of watermark messages can be defined. For example, if the resolution of the HR domain is 8×8, only 64 watermark messages are distinguishable in this system. Since in a watermarking system, we are only interested in identifying the presence of a chirp in the multimedia, a coarse cell division of HRT space, say 32×32, will be enough for our purpose. In this system, if there is a peak in HRT plane, a chirp is detected indicating that the image has been watermarked. If no peak is detected in the HR plane, it means that the image has not been secured.

However, unlike the watermarking technique, a fingerprinting system requires to identify each consumer. Therefore, the system has to be able to assign one chirp to each customer, which means that the resolution of the HRT should match the same number as the client numbers. If the sampling frequency of the watermark chirp is f_s, and the WVD transforms the chirp into a $M \times N$ TF image, the frequency resolution of TFD of the chirp will be $\Omega_f = \frac{f_s}{2M}$, which means that there are Ω_f possible initial and final frequencies. In chirp-based fingerprinting, we should assign a chirp with a pre-defined initial and final frequencies to each user. Therefore, there are $\Omega_f \times \Omega_f$ chirps or fingerprints available that can be assigned to each consumer. Since one of the

challenges in fingerprinting is defining the maximum number of consumers, the HRT plane should be partitioned as fine as possible so that the fingerprinting scheme could contain more consumers. In this study, the length of the chirp is set to 182 bits, and HRT space is divided into 182×182 cells. With these settings, the HRT chirp-based fingerprinting supports about 2^{15} consumers.

DPPT-based Method

The length of the chirp signal has to be selected based on the resolution of the DPPT algorithm. To calculate the required resolution, the performance of the algorithm under noisy conditions has to be considered. If SNR is high, even a short duration of the signal is enough to successfully calculate the signal's phase parameters; however if in low SNR situations, the DPPT algorithm requires more samples of the signal to correctly estimate the parameters. In order to find the optimum length of the chirp, careful attention should be paid in selecting the number of signal samples involved in the estimation process. The DPPT algorithm operates properly when the following inequality is satisfied [53]:

$$M\kappa(M, SNR) \leq \frac{N}{25}. \tag{4.23}$$

where, M is the order of polynomial, N is the number of signal samples and κ is the threshold function and defined as below:

$$\kappa(M, SNR) := \prod_{q=0}^{M-1} \left[\sum_{i=0}^{\binom{M-1}{q}} \binom{\binom{M-1}{q}}{i}^2 i! \left(\frac{1}{SNR}\right)^i \right] - 1. \tag{4.24}$$

In this study, we use DPPT method for linear chirps, so we only consider $M = 2$ case. Fig. 4.12 shows the threshold function κ and the minimum required number of signal samples (N) for different SNRs. As it can be seen in this graph, when the SNR of a given signal decreases, the longer duration of the signal should be used in the DPPT algorithm. Due to the attacks on the watermarked signal, many of the received bits might encounter with errors. Therefore, in this study, we consider cases with low SNRs, say as low as 0db. As demonstrated in Fig. 4.12, for $SNR = 0db$, the algorithm requires minimum 177 bits of the signal. In our implementations, we chose chirp length of 182 bits for watermark and fingerprint messages.

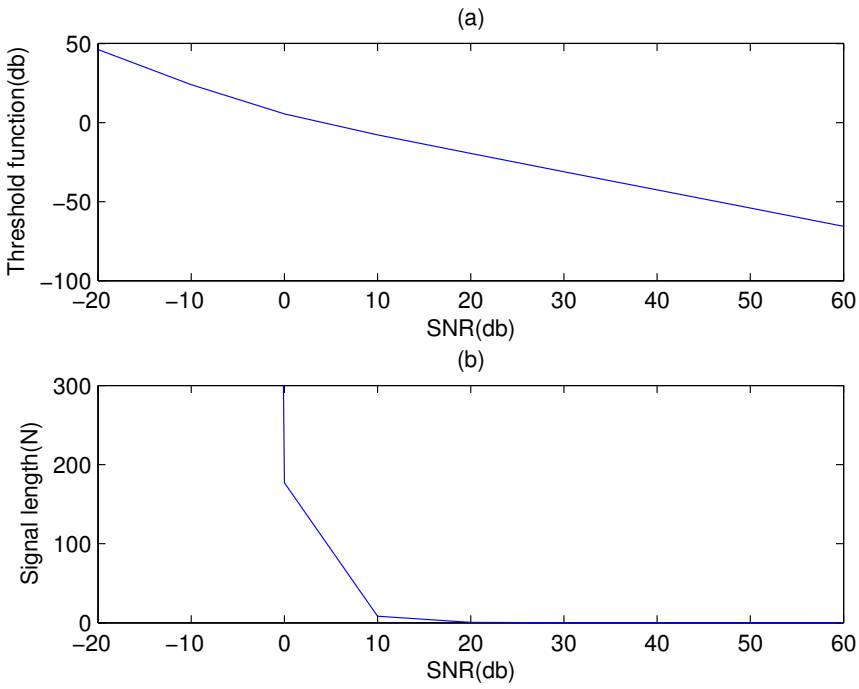


Figure 4.12: (a) Threshold function $\kappa(Q = 2, SNR)$ of the DPPT algorithm vs. SNR. (b) Required number of samples for a linear chirp (Q=2) vs. SNR.

In watermark applications, once the phase parameters are estimated, the following steps are taken in order to verify the embedded watermark. The estimated watermark message $\hat{m}$ is composed as following:

$$\hat{m}_D(n) = \sin\left(\hat{a}_2 n^2 + \hat{a}_1 n + \hat{a}_0\right) \tag{4.25}$$

Then, the correlation between the estimated and the received chirps, $\hat{m}_D$ and $\hat{m}$ respectively, is computed using the following equation:

$$r\left(\hat{m}, \hat{m}_D\right) = \frac{N\sum_{n=1}^{N}\hat{m}(n)\hat{m}_D(n) - \sum_{n=1}^{N}\hat{m}(n)\sum_{n=1}^{N}\hat{m}_D(n)}{\sqrt{N\sum_{n=1}^{N}\hat{m}^2(n) - \left(\sum_{n=1}^{N}\hat{m}(n)\right)^2}\sqrt{N\sum_{n=1}^{N}\hat{m}_D^2(n) - \left(\sum_{n=1}^{N}\hat{m}_D(n)\right)^2}} \tag{4.26}$$

If this correlation is higher than a predefined threshold, the image is considered to be watermarked, otherwise, the image is considered as non-watermarked. In fingerprinting applications, a codebook of all users is defined in a way that each consumer corresponds to a unique chirp in the codebook.

$$\text{Code Book} = \begin{cases} Chirp_{(k)} = \sin\left(a_{k2}n^2 + a_{k1}n + a_{k0}\right) \\ k = 1, ..., K \end{cases} \tag{4.27}$$

where, k refers to the consumer number, and K is the total number of costumers of the fingerprinting system. To fingerprint the document for a specific consumer, say consumer i, the corresponding chirp in the codebook $Chirp_{(i)}$ is embedded into the image. Even using the most robust watermarking techniques, once the fingerprint is extracted, there will be some bit errors in the received chirp $\hat{m}$. Despite the presence of bit error in the received fingerprint, applying the DPPT directly on this signal in time domain is able to obtain a good estimation of the embedded chirp $\hat{m}_D$. To find out that this fingerprint belongs to which customer, we should compare the extracted chirp with all the possible chirps in the codebook to find the best match for the estimated chirp, and recognize consumer i. The process of finding the best match in the codebook is done based on the following methods:

Correlation-based (DPPT[C]): The correlation between the estimated chirp $\hat{m}_D$ with all the possible chirps in the codebook is calculated. The chirp which has the highest correlation coefficient with the estimated chirp is recognized as the embedded chirp, and the corresponding number

of the chirp in the codebook is recognized as the numbers of consumers.

$$\text{Costumer number} = \underset{k=1,...K}{Arg\ \ max}\ r\left(\hat{m}_D, Chirp_{(k)}\right) \tag{4.28}$$

The defined codebook of chirps is not an orthogonal codebook. Therefore, we are expecting a high correlation among the chirps in the codebook that could degrade the preciseness of the fingerprint tracking. In this study, the correlation of the chirps in the codebook is limited to a maximum of 0.93, which results in a fingerprinting capacity of $2^{13.7}$ consumers for a chirp length of 182 bits.

Initial and final frequency-based (DPPT[F]): Since DPPT algorithm provides an estimation of a_0, a_1 and a_2 parameters of the chirp, we can compute the initial and final frequencies of the estimated chirp $\hat{m}_D$ as:

$$\begin{aligned} \hat{f}_0 &= \hat{a}_1 \tfrac{f_s}{2f_r} \\ \hat{f}_1 &= (2N\hat{a}_2 + \hat{a}_1)\, \tfrac{f_s}{2f_r} \end{aligned} \tag{4.29}$$

where, N is the chirp length, and f_r is the minimum possible frequency difference between the initial or final frequencies. This method finds the $Chirp_{(i)}$ in the codebook which has the closest initial and final frequencies with the estimated ones. If the sampling frequency of the watermark chirp is f_s=1 kHz, the initial and final frequencies could be any frequencies in the range of [0 500] Hz. For 182-bit long chirps, the minimum difference between the initial and final frequencies (f_r) in the codebook are defined to be 4 Hz. Hence, we have 125 initial and 125 final frequencies that provide 2^{14} possible fingerprints.

Threshold-based (DPPT[F]-Filter): Before embedding the chirp signal, the energy of the signal in each time is concentrated around the instantaneous frequency of the chirp in that particular time. However, after the chirp is extracted, depending on the number of errors occurred to the chirp, this energy is spread through frequency axis. This scattering of the signal's energy reduces the performance of the DPPT algorithm. Therefore, if we first get a pre-estimation of the chirp signal, and then apply DPPT on those parts of the received signal which correspond to the estimated chirp frequency, the DPPT can result in a more accurate estimation of the chirp's coefficients. Hence, in cases that the bit error rate of the received message is high, first, DPPT[F] obtains an estimation of the initial and final frequencies, $\hat{f}_0$ and $\hat{f}_1$. Then, we take the Fourier transform of the received

chirp signal, $F = \textit{FFT}(\hat{m})$. To remove the scattered energy that degrades the correct estimation of the chirp, we filter the chirp as:

$$F_{filtered}(m,n) = \begin{cases} F(m,n), & \text{if } \frac{2N\pi}{f_s}\left(\hat{f}_0 - 4\right) < m < \frac{2N\pi}{f_s}\left(\hat{f}_1 + 4\right) \\ 0, & \text{otherwise} \end{cases} \tag{4.30}$$

In this filtering process, we are filtering the frequency interval by 4 Hz (the minimum frequency difference in the codebook) from each side to filter the scattered energy. Once, the received chirp is filtered in this frequency range, we take the inverse Fourier transform, and finally, the DPPT[F] algorithm is applied one more time on the filtered chirp to estimate the fingerprint. The modified DPPT[F] technique is referred to as DPPT[F]-filter in the rest of this chapter. Fig. 4.13 shows a case that applying the DPPT[F]-filter technique estimates the embedded chirp successfully, while DPPT[F] fails to extract the correct watermark message.

Forward error correction (FEC)-based Techniques

FEC schemes, or channel codings, are used to protect digital communication by inserting redundant bits into the data. These additional bits contribute in detecting and correcting the errors happened in the data. Due to the similarity between watermarking and communication systems, FEC methods have been commonly used to increase the bit error compensation capacity of fingerprinting techniques. BCH, turbo and repetition codings are the most commonly used FEC schemes in watermarking application [78]. In FEC-based post processing, similar to what we did in the chirp-base fingerprinting technique, a codebook of all the consumers is built up as shown below:

$$\begin{aligned} &\text{Code Book} = \left\{ \textit{Code}_{(k)} = FEC_{encoder}\left(Fingerprint_{(k)}, N\right)\right\}, \\ &Fingerprint_{(k)} = random(L), \\ &k = 1, ..., K \end{aligned} \tag{4.31}$$

where, random(L) creates L-bit random fingerprints for each consumer. $FEC_{encoder}$ encodes the random bits to an N-bit code using one of the FEC schemes, and $\textit{Code}_i$ is embedded into the image of consumer i. In fingerprint tracking process, after the code is extracted from the image, $\hat{m}$, the code is decoded to $\hat{m}_{FEC} = FEC_{decoder}(\hat{m})$. We look up for the match of $\hat{m}_{FEC}$ in the codebook

to recognize the customer of the image:

$$\text{Costumer number} = k \quad if \quad \hat{m}_{FEC} = Fingerprint_{(k)}, \quad k = 1, ...K \tag{4.32}$$

In this study, we utilized BCH and repetition codings as two well-known FEC schemes.

Bose-Chaudhuri-Hocquenghem (BCH) coding: BCH (n, k) is a block coding scheme that segments the data into block of k bits, and transforms each k-bit data block into n-bit block. The $(n - k)$ bits are called redundant bits, and the code rate is k/n. Since in this study our target is comparing different types of post processing methods, all the fingerprint messages used in each method have almost the same number of bits and redundancy rates. n and k for the BCH code that results in a fingerprint message length closer to 180 bits, and gives the highest redundancy rate, are 63 and 7 respectively. BCH $(63, 7)$ encodes a 21-bit fingerprint to a 189-bit long embedded message with $10.7/12$ redundancy rate.

Repetition coding: Repetition coding is a very simple and well-known coding technique. Repetition coding with repetition number of n repeats each bit n times, so results in a redundancy rate of $n/(n + 1)$. We choose $n = 11$ to encode a 15-bit fingerprint to an embedded watermark of 180-bit long and redundancy rate of $11/12$.

4.4.6 Results and Discussion

The data used in this study includes ten different images (512 $\times$ 512) as shown in Fig. 4.14. These images are usually used to evaluate the watermarking techniques. To measure the robustness of the watermark detection algorithms, we perform a series of checkmark benchmark attacks [79]. This checkmark watermark benchmarking was initiated in order to attempt to better evaluate watermarking technologies, and includes a diverse range of common attacks including: data compressions, filtering, sampling, and many others. The spread spectrum watermarking technique with the PN sequence of 100,000 samples and the watermark sampling frequency of 1 kHz is utilized. Once the watermark message is extracted from the data, we apply the watermark detection methods to recover the embedded message. To have a fair comparison of all the watermark detection

Table 4.1: Characteristics of each coding schemes used to code the watermark message

	HRT	DPPT[T]	DPPT[C]	DPPT[F]	REP	BCH(7,63)
Watermark capacity	15	15	13.7	14	15	21
Message length	182	182	182	182	180	189
Redundancy	11/12	11/12	11.08/12	11.09	11/12	10.7/12

Table 4.2: Performance comparison of the HRT and DPPT chirp-based watermarking techniques under Checkmark Benchmark Attacks

Attacks	**Error correction methods**	
	HRT	DPPT[T]
Remodulation(4)	95	90
MAP(6)	100	100
Copy(1)	90	100
Wavelet(10)	100	95
JPEG(12)	100	100
ML(7)	93	80
Filtering(3)	100	90
Resampling(1)	100	100
Color Reduce(2)	85	80
Total Extraction(%)	96	94

techniques, the methods have been set up with almost equal message lengths and redundancy rates. Table 4.1 summarizes these characteristics.

Table 4.2 presents the watermarking detection results for the watermarked images after performing 46 attacks specified in the checkmark benchmark attacks. The first column shows different types of attacks applied on the watermarked image, and the number shows the number of attacks. The number under each column represents the percentage of successful fingerprint detection under each class of attack. As it can be seen in this table, the HRT-based post-processing is more robust than DPPT-based technique; however, DPPT is a faster technique compared to HRT. We also compared the proposed method with some of the techniques in literature. Table 4.3 presents the total watermark detection for the first 5 images in Fig. 4.14 under checkmark benchmark attacks. It is concluded that the chirp-based watermarking offers the most robust watermarking technique.

Table 4.3: Robustness comparison of the proposed method with other methods in the literature under checkmark benchmark attacks.

Watermarking Technology	HRT	DPPT	wang	cox [60]	xia [80]	kim [81]
Total detection(%)	96	94	74	90	84	48

Table 4.4 presents the fingerprint extraction results for the images in 4.1 after performing 46 checkmark benchmark attacks. Table 4.4 shows that DPPT[F]-based method offers higher results when compared to DPPT[C]-based method. In addition, DPPT[F]-based method does not require the long process of correlating the estimated chirp with all chirps in the codebook and is faster than DPPT[C]. In order to improve the performance of DPPT[F] method, as mentioned earlier in this chapter, in cases that the first estimation is not successful, we apply a filter to the received watermark. The low and high cut off frequencies are calculated according to the first DPPT estimation. Then we use DPPT[F] method to reestimate the embedded chirp. The results of this post processing technique is shown in Table 4.4 as DPPT[F]-filter. As presented in this table DPPT[F]-filter has the most robust performance of all the chirp-based post processing.

The fingerprint extraction percentages in Table 4.4 show that DPPT-based method outperforms HRT-based algorithm in most of the attack types with a total detection of 95% compared to 87%. As we mentioned previously, HRT is an effective technique for watermarking; however, in fingerprinting system, the dimensions of the cells in Hough-Radon plane decreases. Therefore, during HRT calculations, the chance of having the peak in a wrong cell, and introducing one of the neighbor cells as the slope of the chirp increases; for example, in a case that the fingerprint is not extracted correctly in 182×182 Hough-Radon plane, when we decrease the cell's number to 32×32 as in watermarking application, the probability of detecting the correct message increases, but decreasing the cell's number decreases the fingerprinting capacity. The other advantage of DPPT-based method is its less complexity compared to HRT-based technique; as we see in Table 4.5, DPPT-based post processing is 55 times faster than HRT-based algorithm, and this makes DPPT more suitable for real-time applications.

Results for BCH and REP codings are presented in Table 4.4. According to this table, the BCH coding results in better extraction results than the ones of REP coding. In addition, as Table 4.1

Table 4.4: Performance comparison of the FEC-based fingerprint extraction schemes and DPPT-based technique under Checkmark Benchmark Attacks

Attacks	**Error correction methods**					
	HRT	DPPT[C]	DPPT[F]	DPPT[F]-filter	REP	BCH(7,63)
Remodulation(4)	60	83	90	95	58	65
MAP(6)	98	100	100	100	97	100
Copy(1)	100	100	100	100	90	100
Wavelet(10)	90	97	96	98	90	92
JPEG(12)	100	100	100	100	100	100
ML(7)	61	73	73	79	57	67
Filtering(3)	100	100	100	100	100	100
Resampling(1)	100	100	100	100	100	100
Color Reduce(2)	70	70	75	75	65	70
Total Extraction(%)	87	92	93	95	85	89

Table 4.5: Order of complexity of each coding schemes used to code the fingerprint.

	DPPT	HRT	REP	BCH(7,63)
Complexity	$O(Nlog_2(N))$	$O(N^2log_2(N)) + O(N_2t)$	$O(N)$	$O(Nlog(N))$
Running time	1.94s	107s	1.86s	1.89s

represents, the BCH coding supports 6 bits more for fingerprint capacity. Table 4.4 shows that DPPT-based method offers better results compared to REP and BCH codings, and shows a higher total extraction than the other methods. Fig. 4.15 graphs the detection results considering BER in the received message. As we see in this figure, DPPT extracts 100% watermark messages successfully up to a BER of 17%, while this value for BCH-based post processing is 16%. Furthermore, the BCH-based technique does not offer any fingerprint extraction for BERs of worse than 22%, while DPPT shows extraction up to BER of 39%. The complexity order and running time of the DPPT-based method shown in Table 4.5 confirms the simplicity and efficiency of the proposed method. The time shown in Table 4.5 is based on Pentium IV, CPU 2.66GHz and 512MB of RAM.

A strong attack that should be considered in a digital fingerprinting is collusion attack, where several users combine their copies of the same content to remove the original fingerprint. The

chirp-based fingerprinting technique that we proposed in this study is robust to collusion attack. The Fig. 4.16 represents a case that three costumers added up their images to remove the original fingerprints. The received message will be the summation of the three chirps:

$$Received\ Chirp = \sum_{i=1}^{3} \sin\Big(a_{i2}n^2 + a_{i1}n + a_{i0}\Big), \tag{4.33}$$

In Fig. 4.16 shows the HRT of the above signal in HR domain. As it could be observed in this figure, there are three peaks in this figure; each peak corresponds to one of the chirps in the extracted watermark. Since each chirp belongs to a customer, we can conclude that three costumers added up their images to remove the fingerprint, and applying the technique of Section 4.3.1, the location of the each peak can be used to recognize each consumer.

4.5 Chapter Summary

This chapter presented a quantification approach of TF signatures with a known structure. The hidden known structure we quantified in this chapter was not part of the signal, and as a matter of fact, it was externally embedded into the signal. We explained HRT and DPPT as the TF and temporal techniques to successfully quantify the time and frequency varying characteristics of the chirps. Their applications were studied as related to multimedia security applications to successfully recover the chirp-based watermark messages. The developed adaptive feature extraction techniques calculated robust TF signatures of the watermark. Then, we used a correlation-based classifier to assign the watermark message to its corresponding user. The performance of each method was evaluated by testing each technique under a set of benchmarked attacks. Our experiments showed that the proposed DPPT method promised a higher detection rate compared to the HRT-based chirp quantification technique, and the FEC schemes such as repetition and BCH codes. The other advantage of the proposed DPPT-based technique was its low computational complexity.

In Chapters 3 and 4, we evidenced the advantage of TF feature quantification when the TF structures of interest were known. The proposed Adaptive TF feature extraction and the DPPT technique successfully quantified and tracked the time and frequency varying structures. However, the discriminative patterns are not always known to us. In fact, this is the case in a majority of

real-world signals where unknown and complex structures have to be detected. In the subsequent chapters, we intend to focus on such feature extraction approaches.

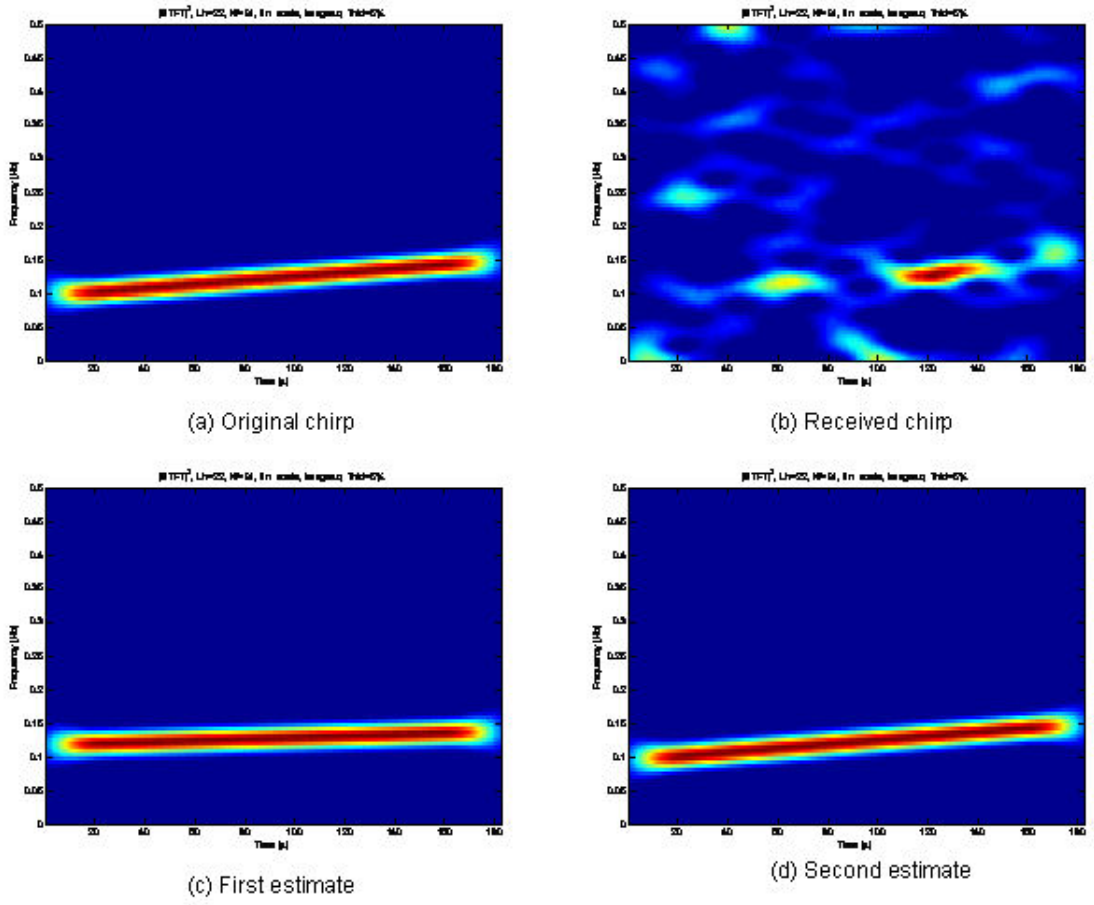


Figure 4.13: Performance improvement of the DPPT-based post processing using DPPT[F]-filter technique. The spectrogram of the embedded chirp, the received chirp at BER of 24% and the first DPPT estimation are shown in (a), (b) and (c) respectively. The initial and final frequencies of the embedded chirp are 75.93 Hz and 117.89 Hz, and the first estimated chirp has 94.44 Hz and 108.76 Hz initial and final frequencies. Since these values have more than 2 Hz difference with the original ones, the watermark extraction is not successful. Part (d) shows the spectogram of the DPPT estimation after filtering the signal; initial and final frequencies are 74.87 Hz and 117.89 Hz. Since the difference is less than 2 Hz with the ones of the embedded chirp, the watermark message is successfully extracted.

Figure 4.14: Test images.

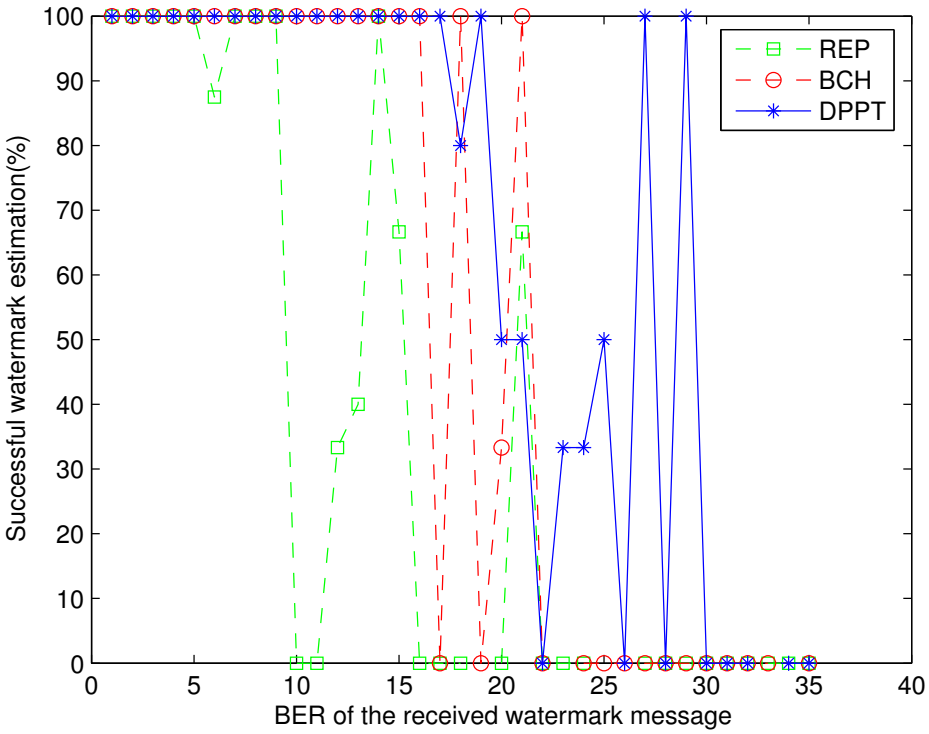


Figure 4.15: Watermark detection under different bit error rates.

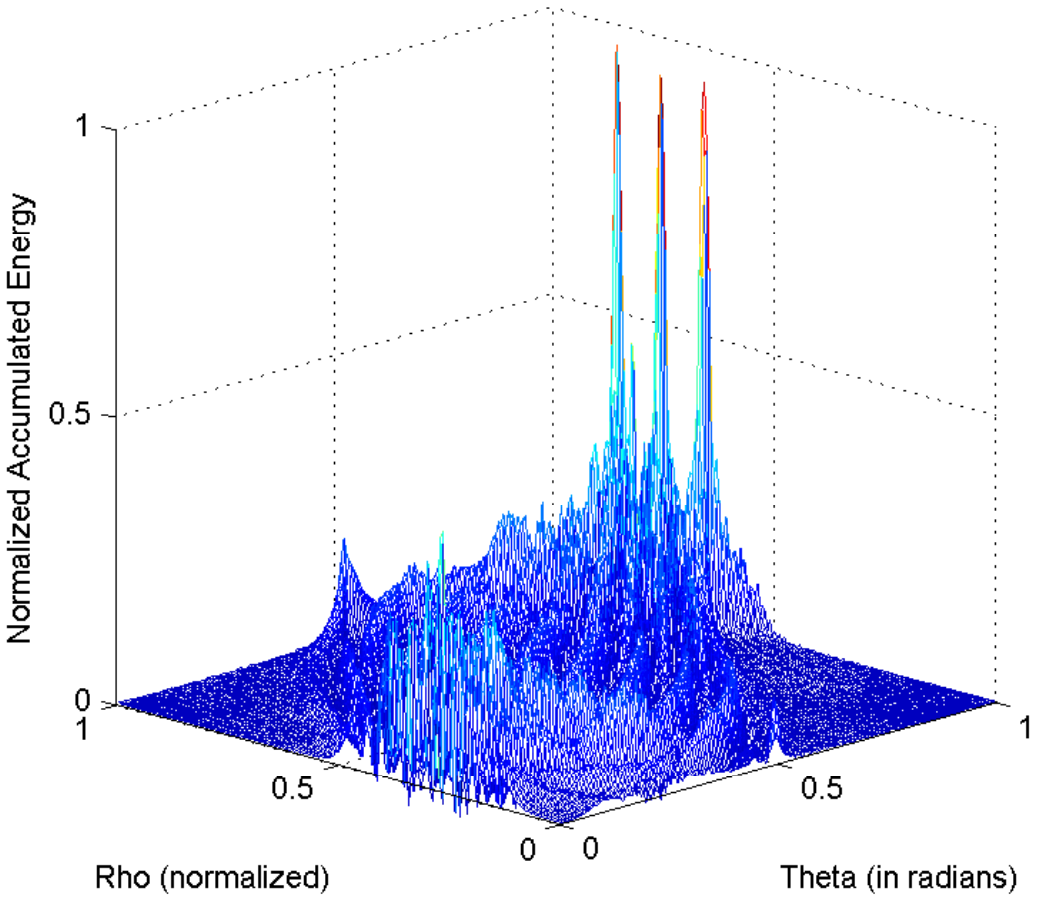


Figure 4.16: The three peaks in HR domain proves that three costumers combined their images.

Chapter 5

TIME-FREQUENCY QUANTIFICATION

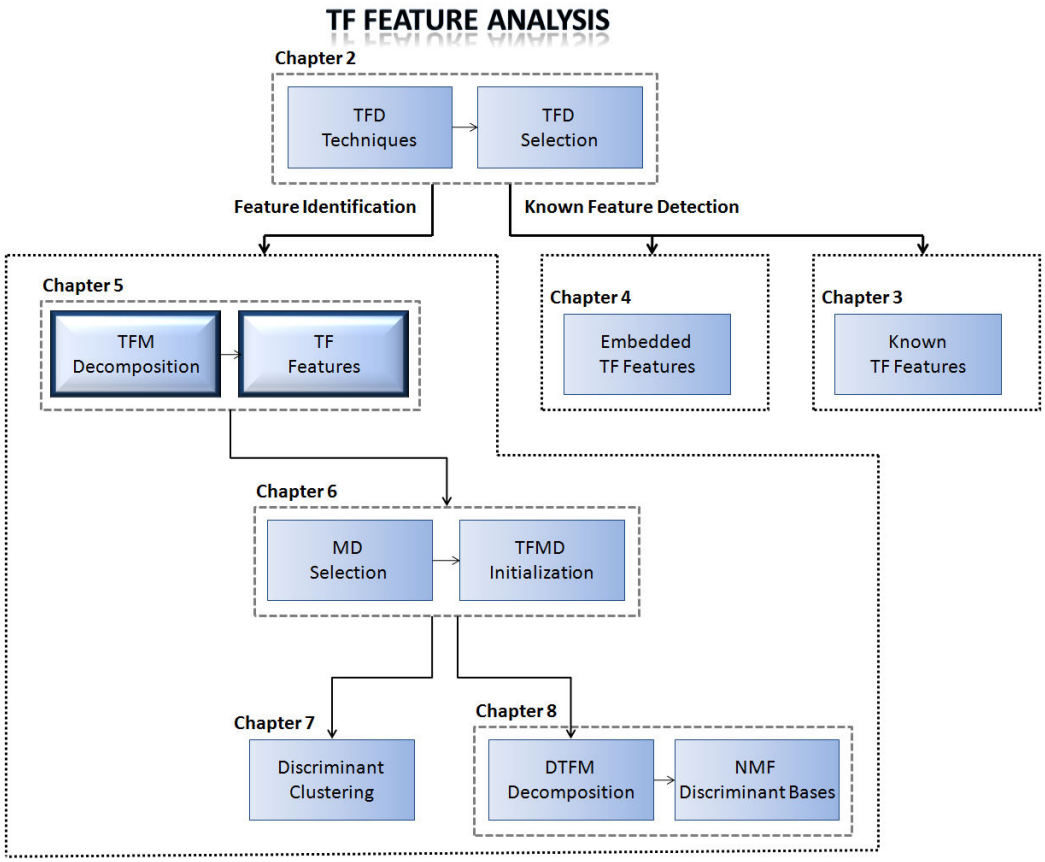


Figure 5.1: Chapter 5 - Developing the TF quantification methodology.

5.1 Motivation

MOTIVATED by the representative characteristics of TF distributions for real-world signals, Chapter 2 selected the TF plane as the first stage of an efficient pattern recognition system for non-stationary signals. In Chapters 3 and 4, we investigated known TF feature detection where Chapter 3 presented an adaptive pattern detection technique verifying the naturally existing structure in a given data, and Chapter 4 developed TF quantification techniques to detect the embedded patterns in a signal. While the proposed approaches promised significant improvements over current feature detection techniques, the developed methodologies had the advantage of knowing the structure being detected, which cannot be assumed true in all the applications. As a matter of fact, the discriminant patterns are usually unknown in real-world pattern detection problems, and our task is to perform the right technique to identify and then detect these unknown and often complex structures. In order to address this demand, the remaining chapters of this dissertation tackle the problem of TF quantification for such unknown and complex structures.

TF representations are potentially strong features for classification of non-stationary signals; however, these representations contain huge amount of information; for example for a 64 ms signal with sampling frequency of 16 kHz and a TFD with resolution of 512×1024 contains 524,288 TF samples. Due to the computational complexity issues, application of such a huge amount of data as features is not efficient. Additionally, employing the entire TF data in the classification stage will adapt the designed system to that specific data, and the system might loss its generality to new signals. Not all the information in the TF plane represent the signal's discriminative region; therefore, to make a TFD suitable for any classification applications, it is essential to reduce the dimensionality of the distribution.

In Chapter 2, our attention focused on identifying an adaptive TFD that was flexible to time or frequency variations in real-world signals. Our inspiration to such a TF plane was to transform the signal into a representation that can result in efficient features as related to pattern recognition. However, representative features will not be obtained from the TFD unless we employ an appropriate TF analysis that reduces its dimension in a non-stationary compatible manner. As the final goal is to analyze long-term non-stationary signals without any stationarity assumptions about

the signal, we focus our attention on developing a new TF quantification method that adaptively breaks down any signal into its stationary parts. Fig. 5.2 highlights the main contributions of this chapter. As it is shown in this figure, the signal processing contains of two parts; in the first part, we construct the TFD of the signal as explained in the previous chapter. After the TFD is constructed, in this chapter we look for a TF analysis that reduces the TF dimensionality by reducing the redundant information in a given TFD. To make this happen, first, we explain the current TF analysis methods and compare the approaches as related to their compatibility to non-stationary quantification. Once the approach is selected, we define the proposed TF analysis in both mathematical and practical points of view. The next stage as shown Fig. 5.2 is feature extraction where we perform a further dimension reduction, and finally we visualize the proposed TF quantification methods through some numerical examples.

5.1.1 Critical Analysis of the State-of-the-Art

Lately, there have been some attempts to reduce the dimensionality of the feature space by removing the redundancy and keeping only the representative parts of the TFD [10, 11, 12, 13, 14, 15, 16, 17], [18, 19]. In general, three notions have been introduced as explained below:

TF Notion 1:

The first TF quantification notion is illustrated in Fig. 5.3(b). In this approach, TFD is constructed as a 2-D probability density function (pdf) of the signal's joint time and frequency behavior, and some representative statistics are extracted from the constructed TF pdf. Although this approach decreases the dimensionality of the TFD to some extend, the signal is still represented with a large number of TF features. To solve this problem, the dimensionality of the TF features have to be decreased before feeding them into the classification stage. One of the first works in this area is [11], in which Loughlin et al. find two dimensional moments of the TF pdf as TF features; however, the drawback of this method is that the quantity of the features is still large. In [12], the same group achieves a further dimension reduction of the feature space by applying principal component analysis (PCA) to the moment features. In a similar work, Kandaswamy et al. [13] decompose

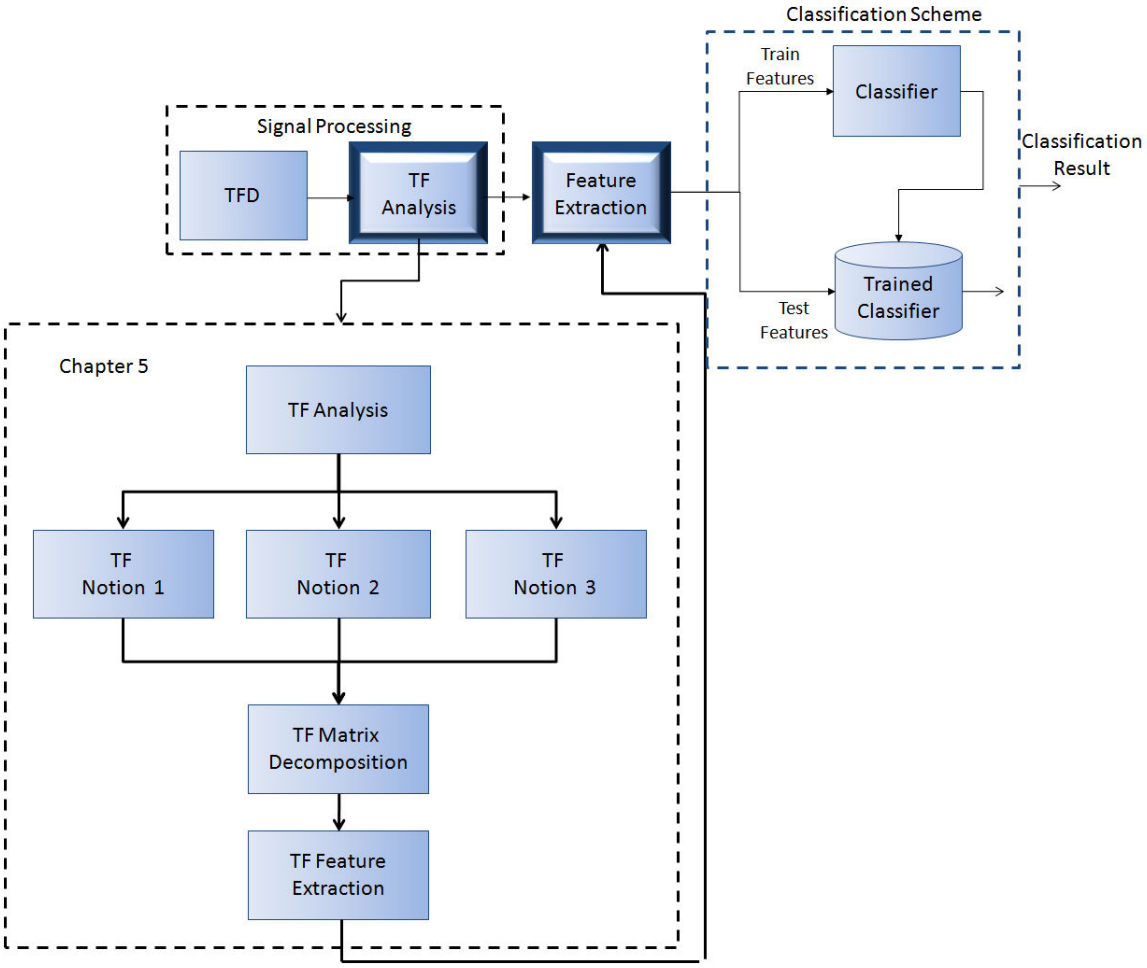


Figure 5.2: Chapter 5 - TF Quantification.

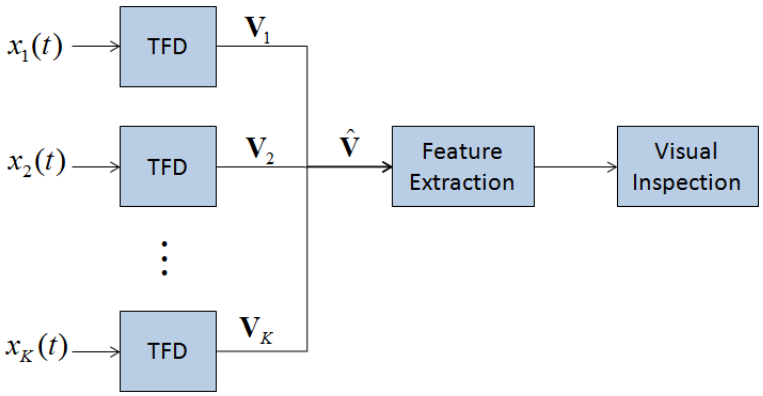
lung sound signals into the frequency subbands using wavelet transform. Instead of applying the TF features directly to classification process, they extract a set of statistical features from the subbands to represent the distribution of the wavelet coefficients. Although these techniques reduce the TF data for classification applications, they basically ruin the localization of the instantaneous features by either averaging the features or by using a dimension reduction technique such as PCA.

TF Notion 2:

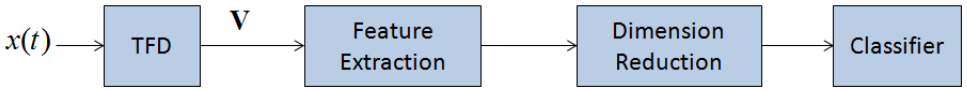
The second TF quantification notion is demonstrated in Fig. 5.3(c). In this methodology, the TFD is interpreted as a matrix (**V**). Then a matrix decomposition (MD) technique is applied to the TF matrix (TFM), **V**, to decompose the TFM into two matrices, **W** and **H**, in a way that $\mathbf{V} \approx \mathbf{WH}$. **W** contains spectral structures, and **H** contains the corresponding temporal location of each spectral structure in the TFM. This notion has been mainly used for separating musical instruments [14]. Recently, the extracted TF vectors ($w_{iM\times 1}, i = 1, ..., r$) from the TFM decomposition have been used for several classification studies. For example, Englehart et al. [15] use this approach for classifying of surface myoelectric signal patterns, Kim et al. [16] use it to extract features for sound classification, and Holzapfel et al [17] use the approach for music classification. While the advantage of the second TF data reduction notion is the absence of any averaging in the temporal domain, the problem existed in this approach is that the dimension of extracted feature vectors are still very large. This is because the length of each feature vector is proportional to the signal's sampling frequency, and as a result they are not very appealing for classification applications.

TF Notion 3:

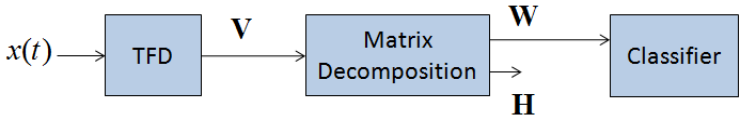
The third TF quantification notion is depicted in Fig. 5.3(d). This method is a combination of the first two notions. As it can be seen in Fig. 5.3(d), the first block uses a MD to reduce the TFM into its spectral and temporal vectors, and the second block decreases the decomposed vectors' dimensionality by considering each vector as a pdf, and extracting the statistical features. TFM feature extraction has not been extensively investigated in literature, and are limited to few works; for example, Groutage and Bennink [20, 21] who apply singular value decomposition (SVD) to the



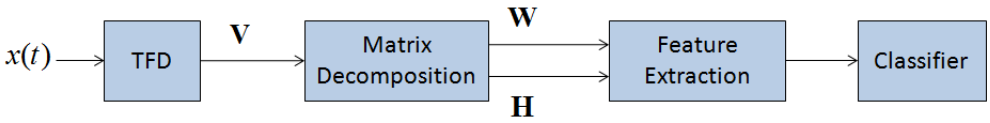
(a) The general block diagram of the TF analysis for the purpose of representation of the event of interest. This approach is mainly performed for multichannel signals, and is limited to visual investigation of the activity of interest.



(b) The general block diagram of the first TF analysis notion. Before employing the TF features in the classification stage, the dimensionality of the features is reduced. This stage may consist of a dimension reduction stage, such as mean or variance.



(c) The general block diagram of the second TF analysis notion. The features that were achieved from the TFM decomposition are directly fed to the classifier. This method preserves the localization of the features, but the dimensionality of the TF features is high.



(d) The general block diagram of the third TF notion which we call TF matrix (TFM) analysis. Decomposed matrices (**W** and **H**) represent the spectral components and their corresponding temporal structures in the TFM, respectively.

Figure 5.3: TF analysis in literature.

TFD constructed from acoustic signatures for the underwater vehicles. The authors showed that the extracted features localize the stair-step pattern of energy density of the transient signal, but, they did not actually use the TFM features for classification purpose. In another work, Boashash et al. [82] propose TF-based features for EEG seizure detection from a single channel EEG waveform using SVD matrix decomposition. This technique extracts the left and right singular vectors (SVs) associated with the TFD of an EEG segment, and finds the histograms extracted from the distribution function formed from the squared-elements of SVs as the TF features. This technique uses MD technique to extract TFM features, but it then diminishes the localization of the TFM features by using the histogram of the SVs as features.

The benefits of the third notion include: i) significant dimension reduction of the TF distribution; and ii) preserving the instantaneous features. While there has been a growing interest in using the second TF quantification notion, the third TF quantification notion for classification applications has not been explicitly studied so far.

5.1.2 Proposed Contribution

In this chapter, we expand the third notion of TF analysis in a way that the presented TF quantification suits pattern recognition applications in terms of both classification accuracy and efficiency. The properties of such a desirable TF quantification are listed in the following criteria:

- **Long-term analysis**. A suitable TF analysis should be capable of analyzing the entire signal rather than blindly segmenting the signal into short frames. Such TF analysis captures the long-term information and the connectivity in a data.

- **Non-stationary compatible**. Although the proposed analysis has to be applicable on long durations of signals, no assumptions can be made regarding the stationarity of a data over the long frames. Therefore, the method has to adaptively deal with the dynamic changes in a given signal.

- **Localized in time and frequency**. As mentioned in Section 1.4, desirable features are needed to be representative of the discriminations between different classes. To make this

happen, the TF analysis should preserve the localization of the TF representation in frequency domain, passing the correct frequency information to the feature extraction stage. Additionally, the proposed analysis is desirable to be localized in time. Localization in time enables us to localize the region of discrimination, which is one more step further than just classifying a signal.

- **Dimension reduction**. Last but not least, the proposed methodology should reduce the dimensionality of the TF plane.

5.2 TF Matrix Decomposition

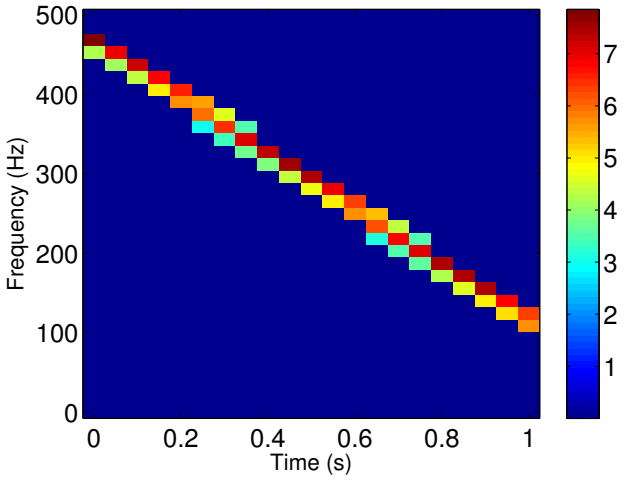
In TFM decomposition methodology, we treat the TF distribution as a matrix, and denote the obtained TF matrix (TFM) of a signal $x(t)$ with $\mathbf{V}_{M\times N}$, where N is the sample length and M is the frequency resolution. For example, let us consider the TFD of a chirp as shown in Fig. 5.4(a). The line in the TFD represents the energy of the chirp with values defined with the color bar beside the TF plot. This TFD has a dimension of 33$\times$ 21. The TFM of this TFD is shown in Fig. 5.4(b). The constructed TFM has the same dimensions as the TFD, i.e. $M = 33$ and $N = 21$, which each entry (m, n) in this TFM has the same content as the (m, n) sample in the TFD. The highlighted elements in the TFM correspond to the chirp line in the TFD. The proposed methodology tends to decrease the dimensionality of the TFM by decomposing the matrix into its significant components without removing any information from the TFD.

5.2.1 Visualization of TFM Decomposition

To better observe the proposed methodology, we display the TF quantification through a synthetic example. In this example, a non-stationary signal is composed of three Gaussian functions modulated with sinusoidal signals as shown in the following equation:

$$x(t) = \sum_{j=1}^{3} x_j(t) = \sum_{j=1}^{3} \alpha_j g\left(\sigma_j, \mu_j\right) \sin\left(2\pi f_j t\right), \tag{5.1}$$

where $g\left(\sigma, \mu\right)$ is a Gaussian with mean μ and variance of σ^2. The mean of this Gaussian function defines where the component is located in time axis, and the variance specifies the temporal du-



(a)

(b)

Figure 5.4: (a) The TFD of a frequency modulated signal. The TF distribution has a dimension of 33× 21. (b) The TFM of the TFD shown in (a). The matrix has 33 rows and 21 columns corresponding to frequency and time resolutions, respectively.

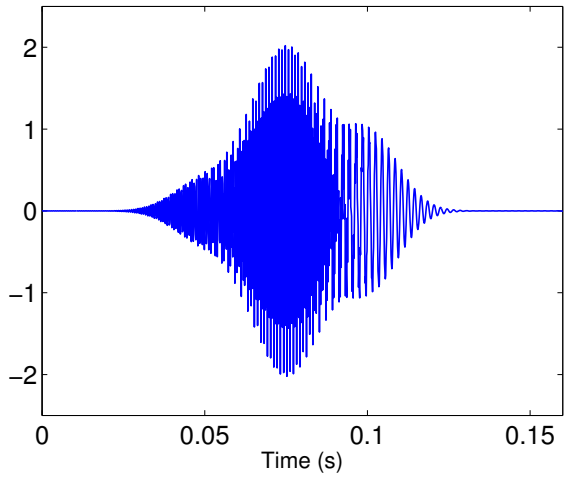
ration of each component. The sine function localizes the component in frequency domain. 160 seconds of such a signal and its TFD are shown in Figs. 5.5(a) and 5.5(b), respectively. The three horizontal lines in the TF plane represent each component. The higher the component is located in this plane, the higher frequency that component has. These three components $(x_1(t), x_2(t), x_3(t))$ are separately shown in Fig. 5.6. The plots on the right are the time representation of each component, and the left ones display the TFDs.

An ideal TF analysis should decompose the TFD in Fig. 5.5(b) into three stationary components by identifying the frequency and temporal location of each component. An example of such an ultimate decomposition is displayed in Fig. 5.7. In this figure, the left plots display three vectors which each represent the frequency structure in a component in the synthetic signal $x(t)$; i.e., the spectral location of each component in frequency axis. The plots on the right side shows three temporal vectors, which each indicates the location of a component in the time domain. Decomposing the TFD into six vectors as shown in Fig. 5.7, not only we reduced the dimensionality of the TFD, but also, we split down the given signal into its multi-components while specifying the temporal locations of each component.

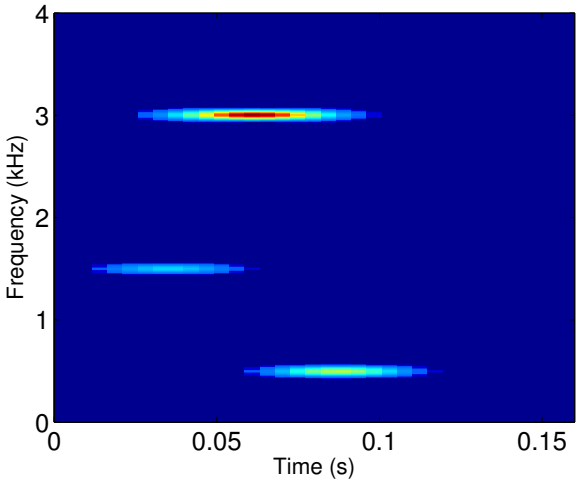
5.2.2 Formulation of TFM Decomposition

In this section, we look for the analytical representation of the TF decomposition methodology explained above. To make such a decomposition happen, first, we treat the TF distribution as a matrix, and denote the obtained TF matrix (TFM) of a signal $x(t)$ with $\mathbf{V}_{M\times N}$, where N is the sample length and M is the number of bins in the frequency axis. For example, the TFD in Fig. 5.5(b) has a dimension of $512\times$ 30; therefore, the constructed TFM has the same dimensions; i.e., $M = 512$ and $N = 30$.

Second, a matrix decomposition (MD) technique is applied to the TFM in such a way that the matrix is decomposed into two sets of vectors denoted with $\{w_1, w_2, ..., w_r\}$ and $\{h_1, h_2, ..., h_r\}$, where w_i represents the frequency structure of each component, and h_i represents the temporal structure of each component. In Fig. 5.7, $\{w_1, w_2, w_3\}$ and $\{h_1, h_2, h_3\}$ are shown on the left and the right side of the figure, respectively. We arrange each vector set into two matrices as defined in



(a)



(b)

Figure 5.5: (a) A synthetic signal is composed of three frequency modulated signal as in Eqn. 5.1; the modulated frequencies are 500 Hz, 1500 Hz, and 3000 Hz. (b) Spectrogram of the same signal with FFT size of 1024 points and Kaiser window with parameter of five, length of 256 samples and 220 samples overlap.

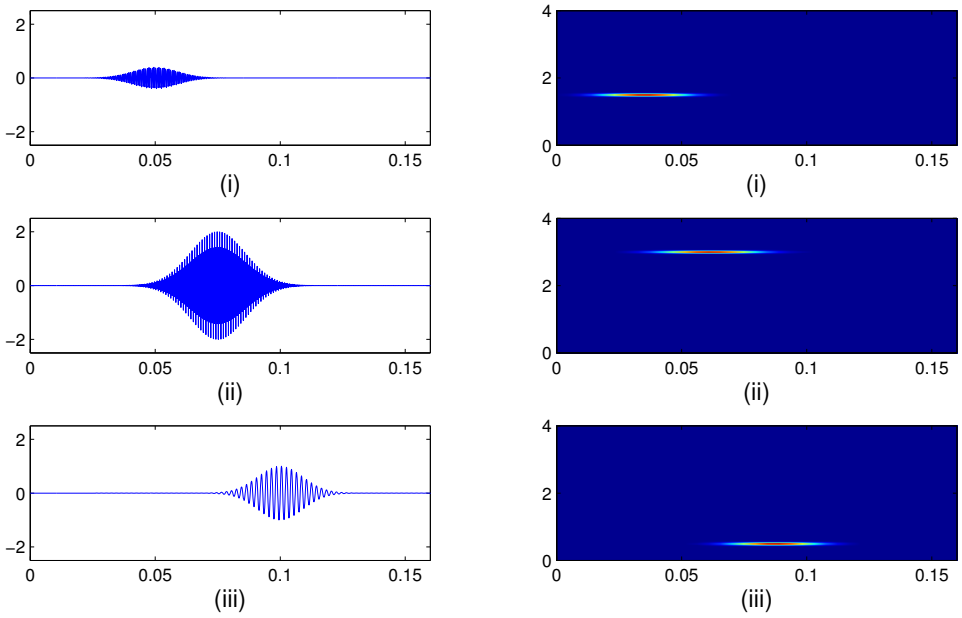


Figure 5.6: The temporal and TFD of each component in the signal of Fig. 5.5 is shown separately. (i) x_1; (ii) x_2; and (iii) x_3.

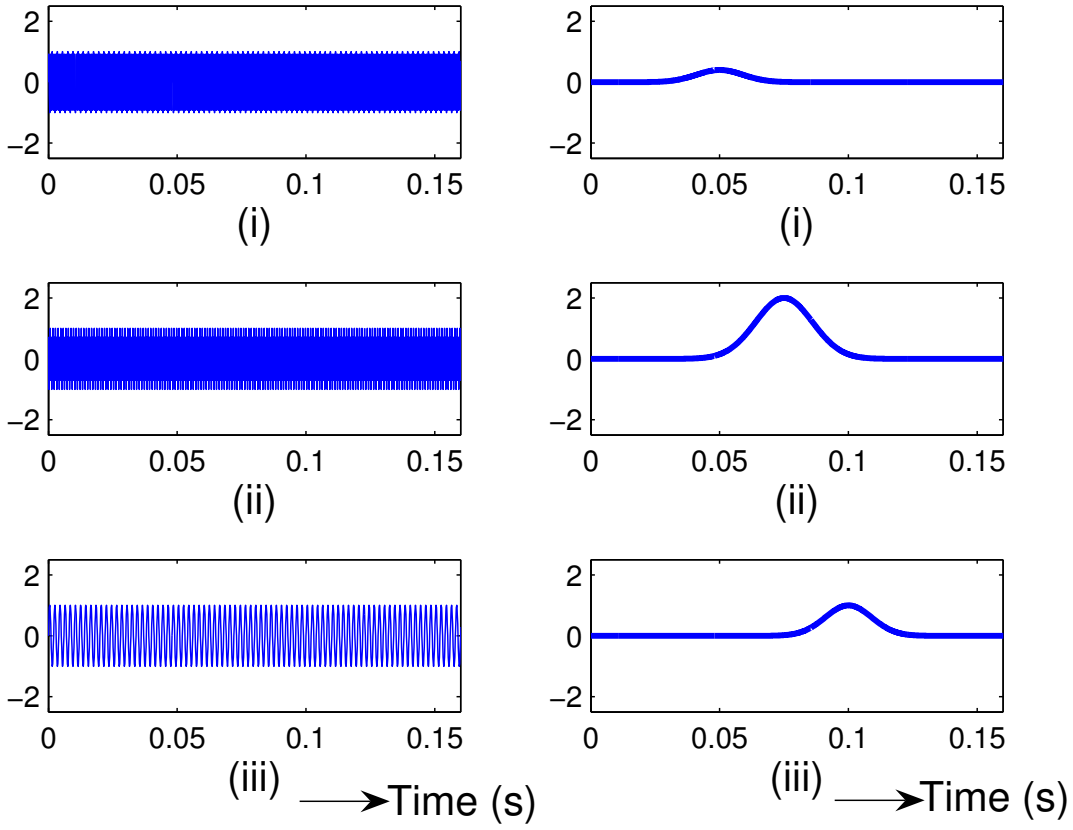


Figure 5.7: Each significant component in the signal of Fig. 5.5 can be represented by two vectors. The left and right plots contain the spectral and temporal structures in each component, respectively. The vertical axes show the time domain in seconds.

the following equation:

$$\mathbf{W}_{M\times r} = [w_1 w_2 \cdots w_r], \tag{5.2}$$

$$\mathbf{H}_{r\times N} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_r \end{bmatrix}, \tag{5.3}$$

and call matrices **W** and **H** as base and coefficient matrices.

Third, we perform the decomposition as shown below:

$$\begin{aligned} \mathbf{V}_{M\times N} &= \mathbf{W}_{M\times r}\mathbf{H}_{r\times N} \\ &= \textstyle\sum_{j=1}^{r} w_j h_j \end{aligned} \tag{5.4}$$

In Eqn. 5.4, MD reduces the TFM (**V**) to the base and coefficient vectors ($\{w_j\}_{j=1,\ldots,r}$ and $\{h_j\}_{j=1,\ldots,r}$, respectively) in a way that the former represents the significant frequency structures in the TFD, and the latter specifies the location of each component in time plane.

5.2.3 Properties

This section analytically proves that the base and coefficient vectors decomposed according to Eqn. 5.4 carry the spectral and temporal information of the signal. The temporal and spectral marginals of the TFM in Eqn. 5.4 (TM and SM, respectively) can be written as follows:

$$\begin{aligned} \text{TM}\left\{\mathbf{V}(f,t)\right\} &= \textstyle\sum_{f=1}^{M} \mathbf{V}(f,t) \\ &= \textstyle\sum_{f=1}^{M}\sum_{j=1}^{r} w_j(f)h_j(t) \\ &= \textstyle\sum_{j=1}^{r}\left\{\sum_{f=1}^{M} w_j(f)\right\} h_j(t) \\ &= \textstyle\sum_{j=1}^{r} \alpha_j h_j(t) \end{aligned} \tag{5.5}$$

and

$$\begin{aligned} \text{SM}\left\{\mathbf{V}(f,t)\right\} &= \textstyle\sum_{t=1}^{N} \mathbf{V}(f,t) \\ &= \textstyle\sum_{t=1}^{N}\sum_{j=1}^{r} w_j(f)h_j(t) \\ &= \textstyle\sum_{j=1}^{r}\left\{\sum_{t=1}^{N} h_j(t)\right\} w_j(f) \\ &= \textstyle\sum_{j=1}^{r} \beta_j w_j(f) \end{aligned} \tag{5.6}$$

where α_j and β_j are the marginals of h_j and w_j vectors. From Eqns. 5.5 and 5.6, TFM decomposition process can be written as given in the following equations:

$$\begin{aligned} \text{TM}\left\{\mathbf{V}(f,t)\right\} &= \alpha_1 h_1(t) + \alpha_2 h_2(t) + ... + \alpha_r h_r(t), \\ \text{SM}\left\{\mathbf{V}(f,t)\right\} &= \beta_1 w_1(f) + \beta_2 w_2(f) + ... + \beta_r w_r(f), \end{aligned} \tag{5.7}$$

As shown in Eqn. 5.7, each sample of frequency and time marginals of the TFM, **V**, is indicated as a linear combination of the base and coefficient vectors. This property demonstrates that base and coefficient vectors carry the spectral and temporal information of the signal, and they can therefore be used for quantification of the signal's TF structure.

5.3 TFM Features

The next stage of TFM quantification shown in Fig. 5.2 is extraction of TF features from the decomposed base and coefficient vectors. Parametrization of the TFM could benefit us in signal classification and feature localization with a significant reduction of the TF features size while preserving the most important information.

Proposed features are $\text{MO}_h, \text{MO}w, \text{S}_h, \text{S}_w, \text{D}_h, \text{D}_w$, and MP. These features are explained as the following:

5.3.1 Joint TF Moments

Joint TF moments of a TFD carry an important information of the TF characteristics of the signal and could be used for classification of time-varying signals [12] and feature identification [20]. For signal $x(t)$, the joint TF moments are given by

$$\left\langle t^{(p)} f^{(q)} \right\rangle = \sum_{t=0}^{N} \sum_{f=0}^{M} t^{(p)} f^{(q)} \mathbf{V}(t,f) \tag{5.8}$$

where, $\mathbf{V}_{M\times N}$ is the TFM of the signal, and q and p are the frequency and time moment orders. The joint TF moments can be used to yield the characteristic function, or moment-generating function, $M(\theta,\tau)$, which can be expanded in a Taylor series given as the following equation:

$$M(\theta,\tau) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{(j\tau)^p (j\theta)^q}{p!q!} \left\langle t^p f^q \right\rangle \tag{5.9}$$

where j in the above equation is the imaginary unit. Although keeping the entire joint moments preserves all of the information in the TFM, it results in a high computational complexity. Therefore, instead of using the entire moments, few joint moments that preserve the major portion of the TF information have to be used as TF features.

If we substitute the TFM, $\mathbf{V}$, in Eqn. 5.8, with its decomposed equivalents as indicated in Eqn. 5.4, the following equation is derived:

$$\left\langle t^{(p)} f^{(q)} \right\rangle = \sum_{t=0}^{N} \sum_{f=0}^{M} t^{(p)} f^{(q)} \sum_{j=1}^{r} w_j(f) h_j(t) \tag{5.10}$$

Since $t^{(p)}$ and $f^{(q)}$ are always non-negative terms, if the term, $w_j(f)h_j(t)$, for $j = 1, ..., r$ is also positive, then Eqn. 5.10 can be re-written as shown below:

$$\begin{aligned} \left\langle t^{(p)} f^{(q)} \right\rangle &= \textstyle\sum_{j=1}^{r} \sum_{t=0}^{N} t^{(p)} h_j(t) \sum_{f=0}^{N} f^{(q)} w_j(f), \\ &= \textstyle\sum_{j=1}^{r} \left\langle t_j^{(p)} \right\rangle \left\langle f_j^{(q)} \right\rangle \end{aligned} \tag{5.11}$$

where $\left\langle f_j^{(q)} \right\rangle$ and $\left\langle t_j^{(p)} \right\rangle$ are the moments of the base and coefficient j respectively. From Eqns. 5.9 and 5.11, one can conclude that instead of calculating the joint TF moments, we can easily determine moments of base and coefficient vectors separately, and use them as the TF features of a given signal.

We denote the pth temporal and qth spectral moments with $\mathrm{MO}_h^{(p)}$ and $\mathrm{MO}_w^{(q)}$, respectively, and compute them as the following equations:

$$\begin{aligned} \mathrm{MO}_{h_j}^{(p)} &= \mathrm{Log}_{10} \textstyle\sum_{t=0}^{N} \left(t - \mu_{h_j}\right)^{(p)} h_j(t), \\ \mathrm{MO}_{w_j}^{(q)} &= \mathrm{Log}_{10} \textstyle\sum_{f=0}^{M} \left(f - \mu_{w_j}\right)^{(q)} w_j(f), \end{aligned} \tag{5.12}$$

where μ_{h_j} and μ_{w_j} are averages of the jth coefficient and base vectors, respectively.

5.3.2 Sparsity

S_{h_j} and S_{w_j} are the sparsity of coefficient and base vectors, respectively. This feature helps to distinguish between transient and continuous components. Several sparseness measures have been

proposed and used in literature. We use a sparsity function as follows:

$$\begin{aligned} \mathrm{S}_{h_j} &= \mathrm{Log}_{10} \frac{\sqrt{N} - \left(\sum_{t=0}^{N} h_j(t)\right) / \sqrt{\sum_{t=0}^{N} h_j^2(t)}}{\sqrt{N} - 1}, \\ \mathrm{S}_{w_j} &= \mathrm{Log}_{10} \frac{\sqrt{M} - \left(\sum_{m=0}^{M} w_j(f)\right) / \sqrt{\sum_{f=0}^{M} w_j^2(f)}}{\sqrt{M} - 1}, \end{aligned} \tag{5.13}$$

The sparsity is zero if and only if a vector contains a single non-zero component, and is negative infinity if and only if all the components are equal. The sparsity measure in Eqn. 5.13 has been used for applications such as developing a matrix decomposition with more part-based properties [83]; however, it has never been used for feature extraction application.

5.3.3 Discontinuity

D_h and D_w represent the discontinuities and abrupt changes in each vector. These features are calculated as follows:

$$\mathrm{D}_{h_j} = \mathrm{Log}_{10} \sum_{t=0}^{N-1} h_j'(t)^2, \tag{5.14}$$

$$\mathrm{D}_{w_j} = \mathrm{Log}_{10} \sum_{f=0}^{M-1} w_j'(f)^2, \tag{5.15}$$

where h_j' and w_j' are derivative of coefficient and base vectors, respectively, as defined in the following equations:

$$h_j'(t) = h_j(t+1) - h_j(t), \tag{5.16}$$

$$t = 0, ..., N-1 \tag{5.17}$$

and (5.18)

$$w_j'(f) = w_j(f+1) - w_j(f), \tag{5.19}$$

$$f = 0, ..., M-1 \tag{5.20}$$

D_h and D_w capture the discontinuities and abrupt changes in coefficient and base vectors, respectively. A vector with a smaller value of discontinuity feature is smoother compared to a vector with a larger discontinuity feature.

5.3.4 Coherency

MP is the Matching Pursuit Feature. Using I iterations of MP as shown in Eqns. 2.11 and 2.12 of Chapter 2, we project a signal into a linear combination of Gaussian functions $g_{\gamma_i}(t)$. The amount of signal energy projected at each iteration depends on the signal structure. A signal with coherent structure needs less number of iterations, while a non-coherent structured signal takes more iterations to complete the decomposition. The term of coherency has been also used in time series and spectral analysis; it is used to describe the strength of mutual relationship between two-series which could depend on the possible time delay, scaling, or even filtering. The Coherency feature that we propose in this dissertation differs from the coherency explained above, and cannot be extracted from the time series. We adopted the name of coherency to characterize the inherent correlation in the signals' structures. Fig. 5.8 shows the projection energy at each iteration for one coherent and one non-coherent segment. The non-coherent segment belongs to 87 ms of an aircraft audio sample, and the coherent one is an 87 ms segment from a piano audio sample. As it can be observed in these figures, in the case of the coherent signal, the major portion of the signal energy is projected before 20 iterations; however, for the non-coherent segment, still after 100 iterations, the signal is not completely projected. We use this property to propose MP features from a signal.

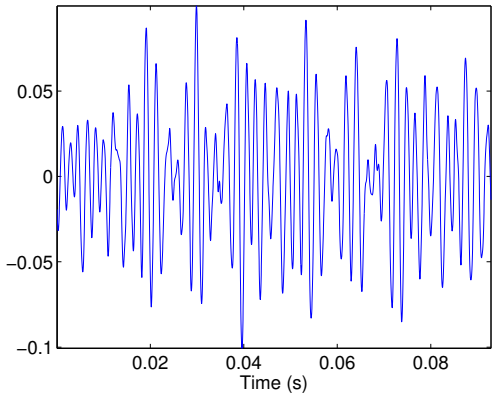
In order to calculate MP feature in a way that it discriminates coherent signals from non-coherent ones, and is independent from signal's energy, we performed the following steps. First, we find the projection energy difference between iteration $i+1$ and i:

$$\begin{aligned} &d(i) = \left(a_{\gamma_{i+1}} - a_{\gamma_i}\right)/E, \\ &i = 1, ..., I-1 \end{aligned} \tag{5.21}$$

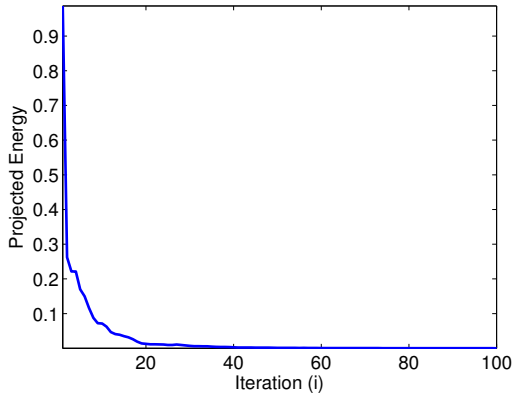
where a_{γ_i} is the MP projection energy ratio at each iteration as shown in Eqn. 2.11, and E is energy of the signal, $x(t)$. Next, we define $L(i)$ as follows:

$$\begin{aligned} &L(i) = \textstyle\sum_{k=1}^{i} d(k), \\ &i = 1, ..., I-1 \end{aligned} \tag{5.22}$$

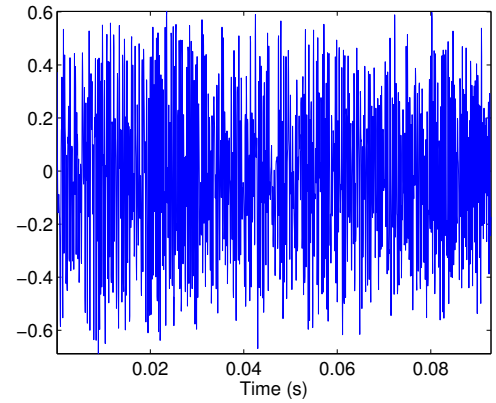
Fig. 5.9 shows L coefficients which represents the normalized coefficients of the energy projection shown in Fig. 5.8. As it can be seen in this figure, L keeps the trend of the energy projection coefficients (a_γ), but it is normalized and it is independent of the signal's energy. Finally, normalized



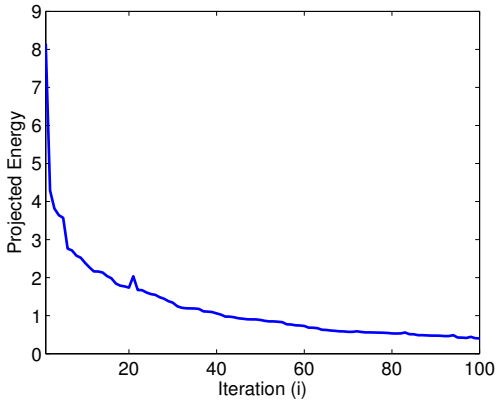
(a) Piano segment



(b) MP energy decomposition for Piano segment



(c) Aircraft segment



(d) MP energy decomposition for aircraft segment

Figure 5.8: (a) one segment from a piano audio sample; (b) the matching pursuit energy decomposition (a_γ) of the piano segment in (a); (c) one segment from an aircraft sample; and (d) the matching pursuit energy decomposition (a_γ) of the aircraft segment in (c).

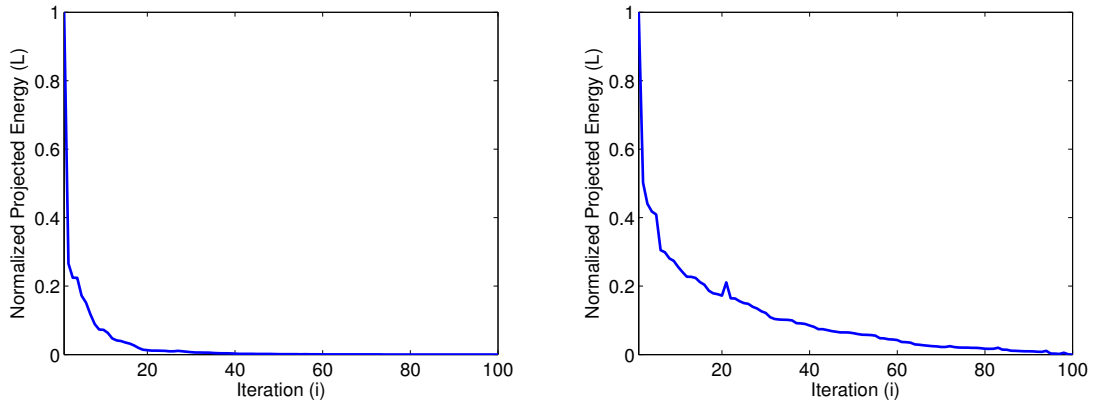
coefficients ($L(i)$) are used to calculate MP feature as follows:

$$MP = \text{Log}_{10} \sum_{i=1}^{I-1} L(i), \tag{5.23}$$

Combining Eqns. 5.21 and 5.22, the above MP feature can also be represented as follows:

$$MP = \text{Log}_{10} \left\{ \frac{\sum_{i=2}^{I} a_{\gamma_i} - (I-1)a_{\gamma_1}}{E} \right\}, \tag{5.24}$$

The MP feature for piano and aircraft signal is calculated as 2.9 and 10.6, respectively. As it is expected, MP feature is higher for the non-coherent segment, and it is lower for the coherent segment.



(a) (b)

Figure 5.9: Normalized energy projection (L), calculated using Eqn. 5.22, is shown for the piano and aircraft segments in Figs. 5.8(b) and 5.8(d), respectively. (a) represents the piano segment with MP feature of 2.9; and (b) represents the aircraft segment with MP feature of 10.6.

In this section, several TF features were extracted from the significant TF spectral and temporal components of a given signal. The extracted features are novel in the sense they intelligently characterize the long-term information of the signal. They are unique to the proposed TFM decomposition method and difficult to capture with any other methods. The summary of the extracted TF

features are listed in Table 5.1. This table also demonstrates the properties of TF features based on TF notions 1 and 2 as explained earlier in this chapter.

Table 5.1: Desirable Properties of the Extracted TF Features

Property	Long-term Analysis	Non-stationary Compatible	Localized In Time	Localized In Frequency	Dimension Reduction
Joint TF Moments (**H**)	×	×	×		×
Joint TF Moments (**W**)	×	×		×	×
Sparsity (**H**)	×	×	×		×
Sparsity (**W**)	×	×		×	×
Discontinuity (**H**)	×	×	×		×
Discontinuity (**W**)	×	×		×	×
Coherency	×	×			×
TF Notion 1		×			×
TF Notion 2	×	×			

5.4 Experiment: Synthetic Signal and Pathological Speech

The application of TF quantification is demonstrated through several examples. First, TFM decomposition is demonstrated using two examples: a synthetic signal and a real world signal. Second, TFM feature extraction is displayed.

5.4.1 Visualization of TFM Decomposition

Synthetic Signal

Figure 5.10 demonstrates TFM decomposition of a synthetic signal. A synthetic signal $x(t)$ is constructed using seven frequency modulated signals similar to the signal in Eqn. 5.1. Fig. 5.10(a) represents the synthetic signal in time, and Fig. 5.10(b) shows the corresponding TFM. TFM is constructed using spectrogram method, FFT size of 1024 points and Kaiser window with parameter of 5, length of 256 samples and 220 samples overlap. A matrix decomposition (MD) with decomposition order of five ($r = 5$) is applied to the TFM, and the decomposed matrices are depicted

in Figs. 5.10(c) and (d). As evident in these figures, each column of the base matrix characterizes a component in the TFM, and each corresponding row of the coefficient matrix illustrates the temporal location of that component in TFM. Fig. 5.11 shows the decomposed TF matrices. This example showed that rather than blindly segmenting a signal into smaller frames, TFM decomposition divided the non-stationary signal into components with similar spectral characteristics.

Pathological Speech Example

An example of TFM decomposition for a 470 ms segment of a pathological speech signal from the Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database [84] is illustrated in Fig. 5.12. In this figure, TFM is constructed using spectrogram as explained in the synthetic example. Figs. 5.12(a) and 5.12(b) show the speech signal in time and TFM representations, respectively. The first 100 ms of the signal belongs to an unvoiced sound and the rest of the signal belongs to three voiced sounds. A MD is applied to the above TFM and six components were decomposed. Figs. 5.12(c) and 5.12(d) illustrate the decomposed bases and the corresponding coefficients of each base, respectively. It can be seen in the figures that each column of the base matrix characterizes a component in the TFM, and each corresponding row of the coefficient matrix illustrates the temporal location of that component in the TFM. For example, the first column in the base matrix represents the spectral characteristics of the first formant in the voiced part of the speech signal. Correspondingly, the coefficient vector verifies the presence of this formant during the same voiced speech. The third base vector in Fig. 5.12 (c) represents the frequency structure of the unvoiced part of the speech, and its corresponding coefficient vector shows the presence of this spectral structure over the first 100 ms of the speech signal.

The above examples demonstrated that TFM decomposition considerably reduces the original TFM into only r pairs of base and coefficient vectors, yet it preserves almost all the significant information in the TFM. Therefore, instead of traditionally assuming the stationarity of the signal over 20-23 ms, the above method adaptively decomposes the TFM into intervals with similar spectral characteristics.

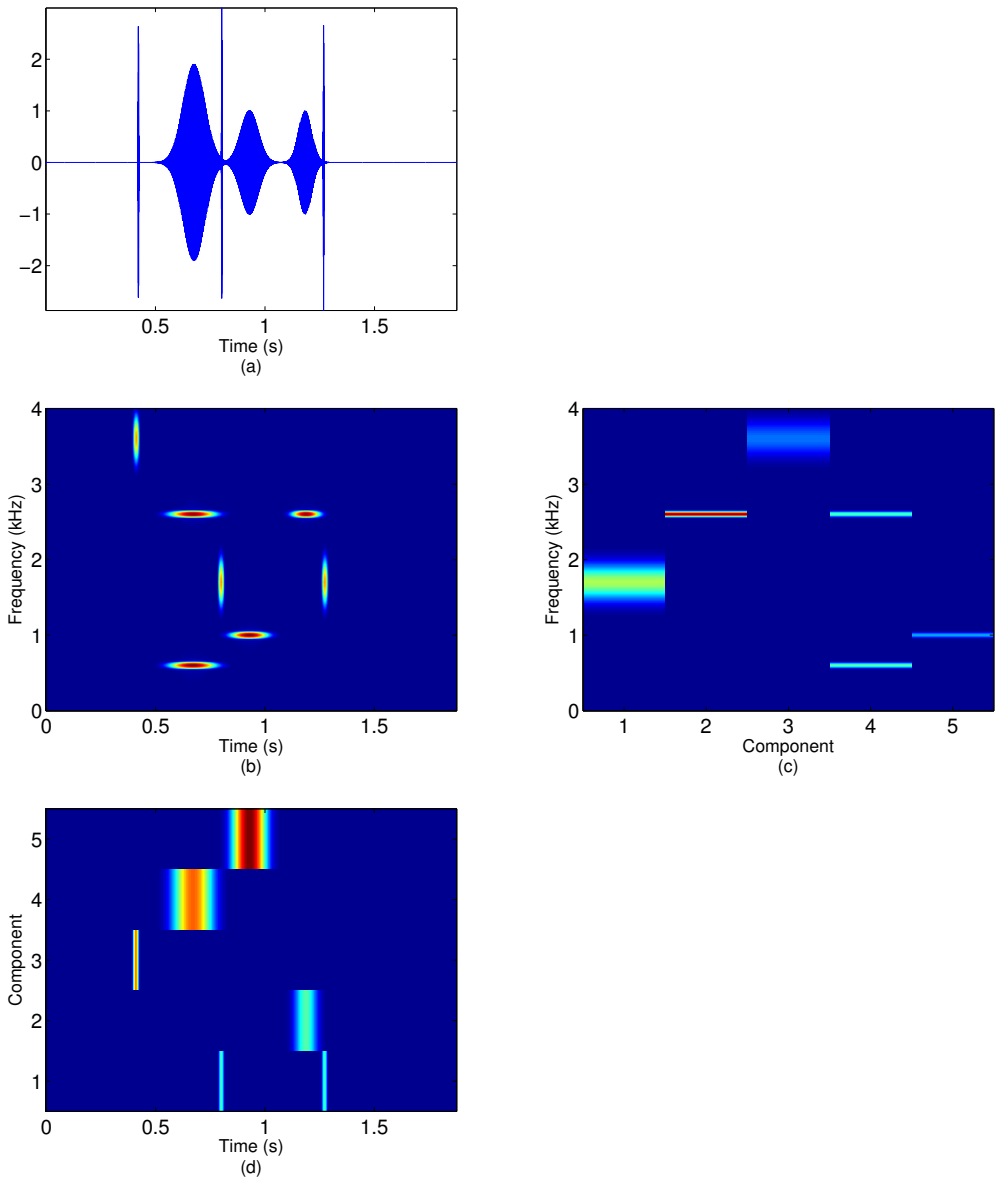


Figure 5.10: Eqn. 5.1 is used to generate a synthetic signal. (α, σ, μ, a) for each component from 1 to 7 is as following: (3,0.001,0.42,2π 3600), (1,0.05,0.68,2π 2600), (1,0.05,0.68,2π 600), (3,0.008,0.8,2π 1700), (3,0.008,1.27,2π 1700), (1,0.04,0.93,2π 1000), (1,0.03,1.18,2π 2600). (a) The synthetic signal in time domain. (b) The TFM of the constructed signal. (c) The decomposed base matrix; each column of the base matrix represents a TF character in TFM. (d) The coefficient matrix; each row shows a coefficient vector.

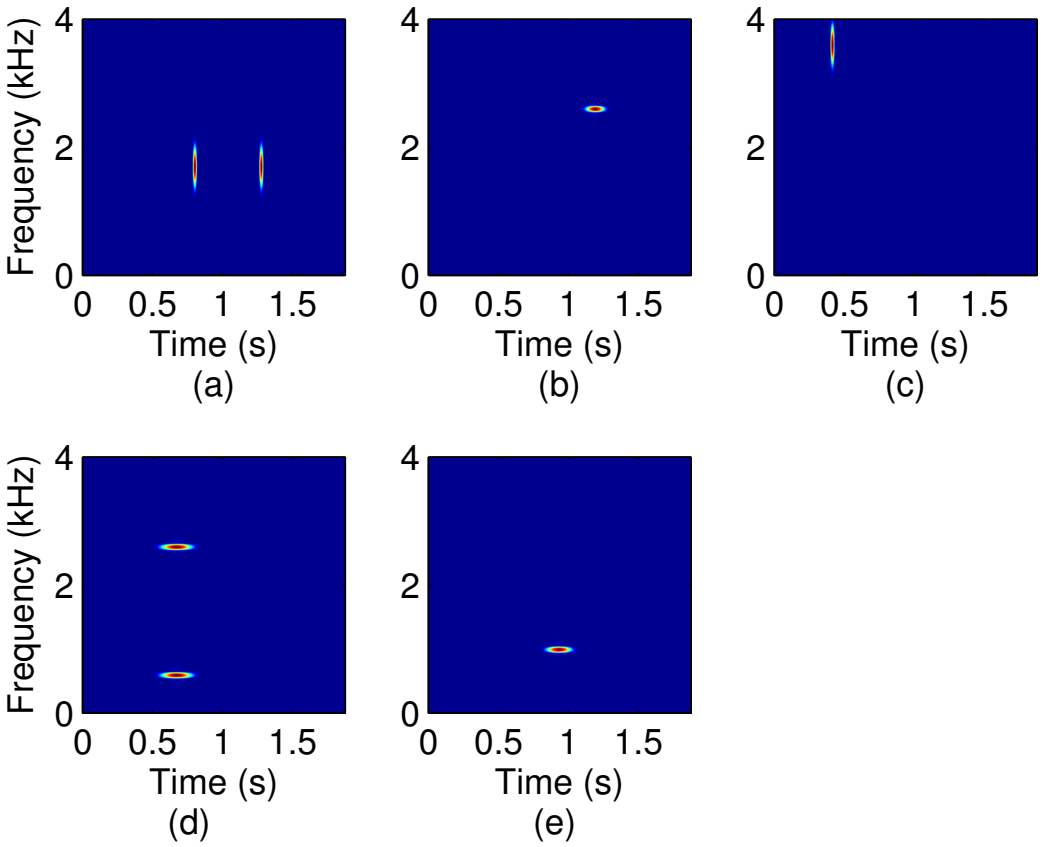


Figure 5.11: The decomposed matrices from the TFM in Figure 5.10 are displayed. (a) to (e) correspond to the decomposed bases 1 to 5 in Figure 5.10(c), respectively.

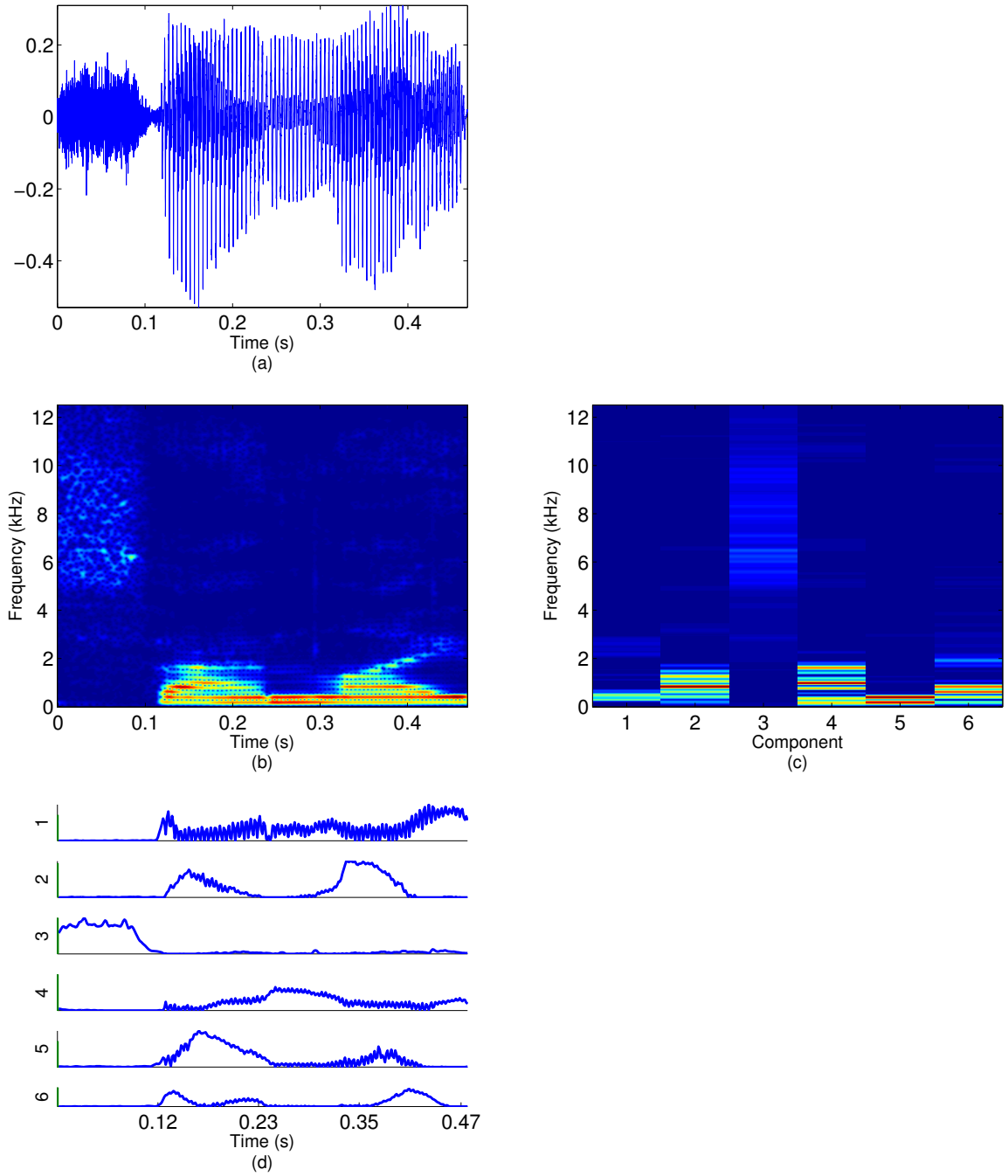


Figure 5.12: A 470 ms segment of a pathological speech signal with sampling frequency of 25 kHz and quantization resolution of 16 bits/sample is analyzed using TFM decomposition method. (a) The signal in time domain. (b) The TFM of the pathological segment. (c) The decomposed spectral vectors (bases). (d) The decomposed temporal vectors (coefficients).

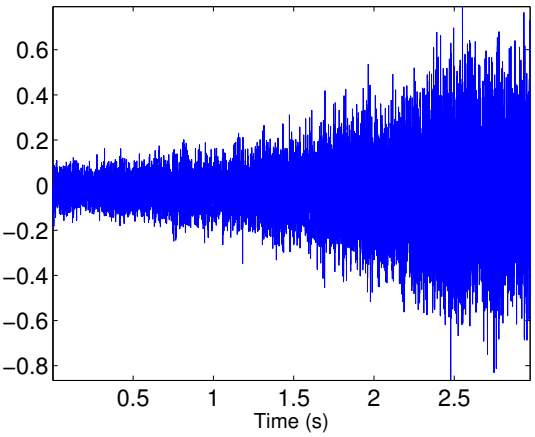
5.4.2 Visualization of TF Features

In the following section, we demonstrate the TF features obtained from two signals with different properties, and investigate if the extracted features from different signals are separated in the feature domain. Figs. 5.13 and 5.14 show the TFM decomposition process for a plane and a piano signal, respectively. The decomposition matrices show the spectral and temporal location of significant components in each signal. The feature plane shown in Fig. 5.15 demonstrates the feature vectors that are extracted from the aircraft and the piano signals. It can be observed in the feature space that the feature vectors of aircraft are located far away from the piano features. As it was expected, S_W and D_H in the piano features are smaller than the aircraft features. This is because the piano's spectral components in **W** are less sparse than the aircraft's components. Also, piano's coefficient vectors are smoother than the aircraft's coefficient vectors.

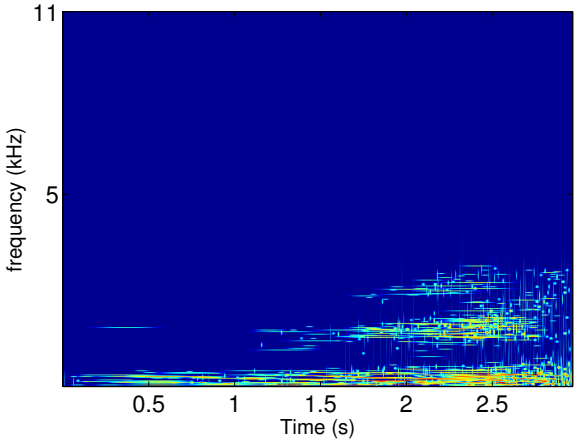
5.5 Chapter Summary

Fig. 5.16 displays the contribution flowchart, and the highlighted blocks show the progress of the work in this chapter. This chapter presented the time-frequency matrix (TFM) quantification based on TFM decomposition and new TF features, which intelligently characterized the long-term information in a non-stationary signal. TFM decomposition is a window-less approach that was applied to the entire data without any need to segment the data into short durations. The proposed TF features preserved the time and frequency localization of a given signal, and provided a significant low-dimensional and yet powerful quantification tool for real-world signals. The new TF features were unique to the proposed TFM quantification framework, which are difficult to be extracted by other methods. The performance of the proposed TF quantification methodology was demonstrated through some synthetic and real signals.

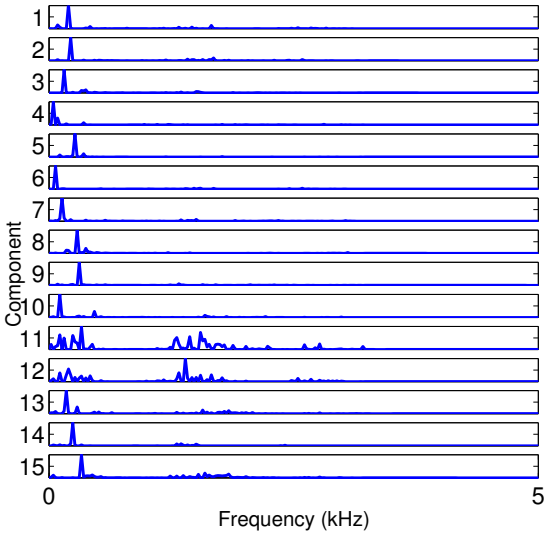
As can be seen in Fig. 5.16, the next stage is calibration of the proposed TF quantification framework. The next chapter investigates the most suitable matrix decomposition technique as related to TFM quantification. A complete evaluation of the TFM quantification method using the selected matrix decomposition technique will be presented in the next chapter.



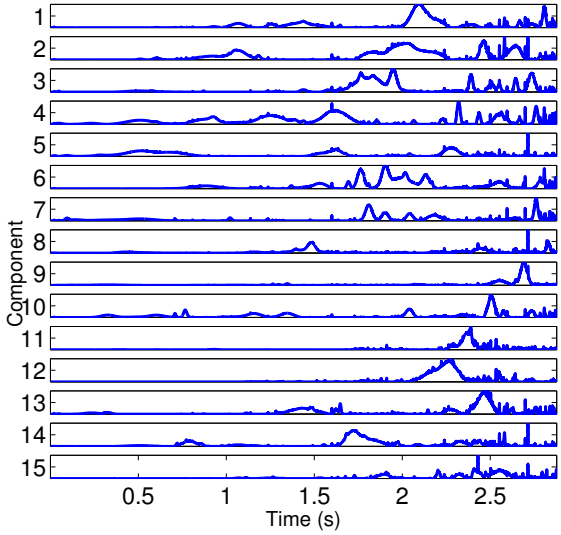
(a) Time representation



(b) MP-TFD representation

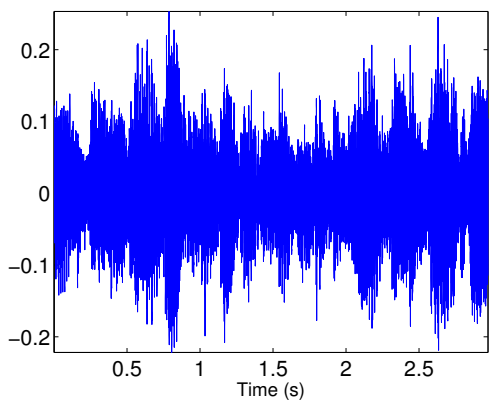


(c) Base vectors decomposed using non-negative matrix factorization (NMF) technique

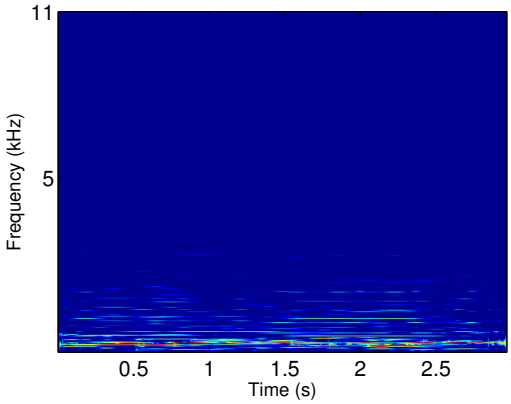


(d) Coefficient vectors decomposed using NMF technique

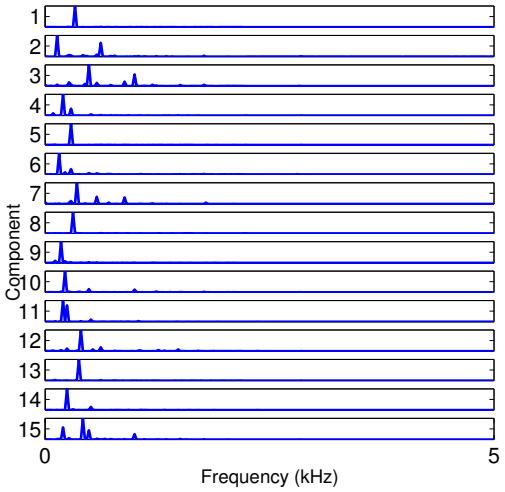
Figure 5.13: (a) and (b) show a segment that belongs to an aircraft signal in time and MP-TFD representations, respectively. Applying NMF to the TFM in (b), we extract 15 base and coefficient vectors which are depicted in (c) and (d), respectively.



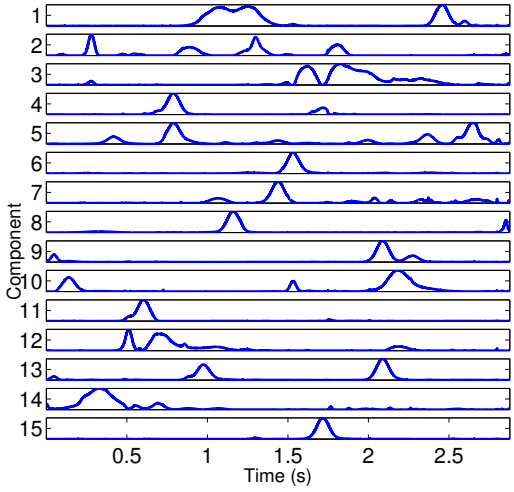
(a) Time representation



(b) MP-TFD representation



(c) Base vectors decomposed using NMF technique



(d) Coefficient vectors decomposed using NMF technique

Figure 5.14: (a) and (b) show a segment that belongs to a piano signal in time and MP-TFD representations, respectively. Applying NMF to the TFM in (b), we extract 15 base and coefficient vectors which are depicted in (c) and (d), respectively.

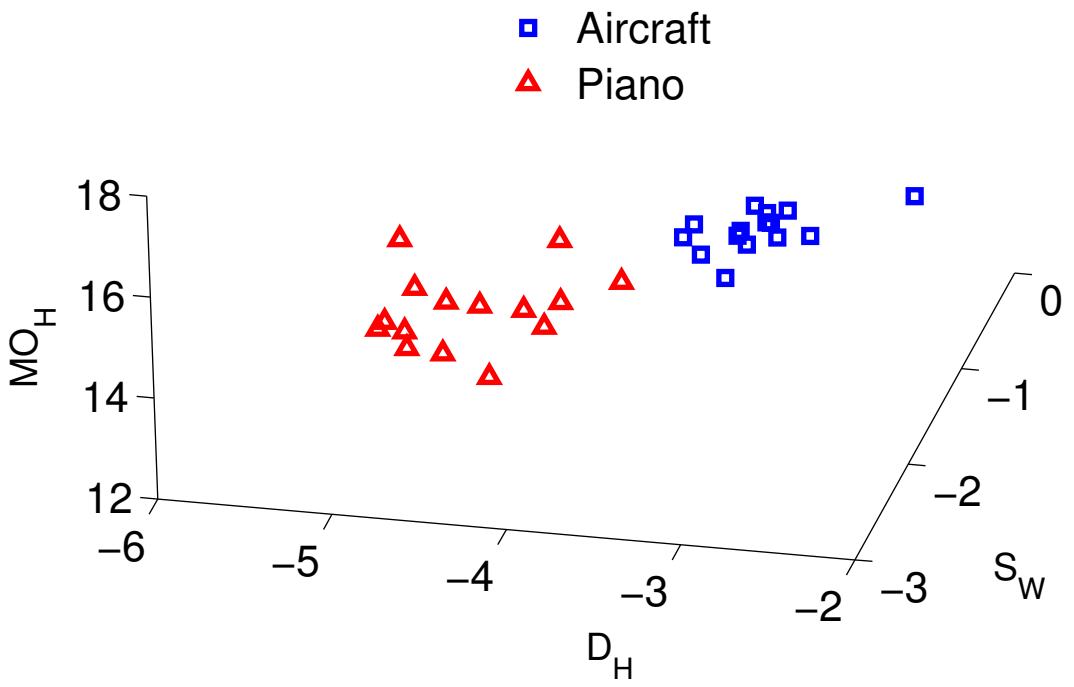


Figure 5.15: This figure represents the aircraft and piano segments in the feature plane. Three features of the feature vectors are shown in this figure. MO_H, D_H, and S_W represent the second central moment of coefficient vectors in **H**, the derivative of coefficient vectors in **H**, and the sparsity of base vectors in **W**, respectively. As it can be observed from the feature domain, the feature vectors from aircraft and piano are separate from each other.

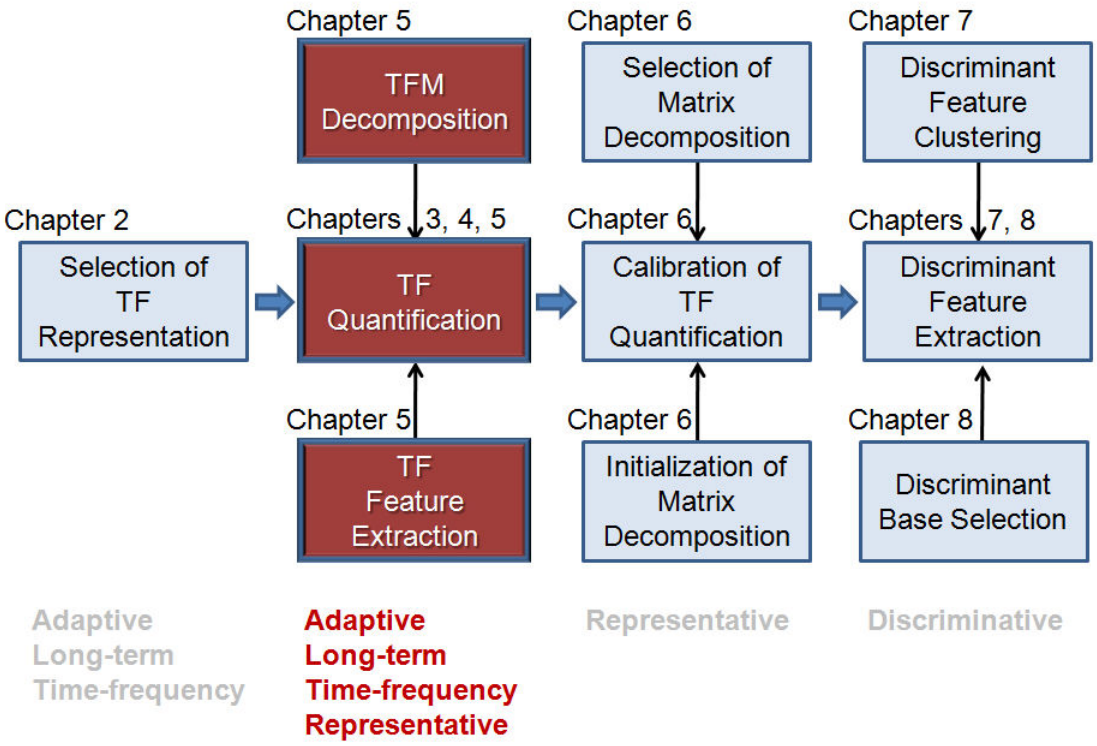


Figure 5.16: Flowchart of the proposed contributions.

Chapter 6

MATRIX DECOMPOSITION ANALYSIS

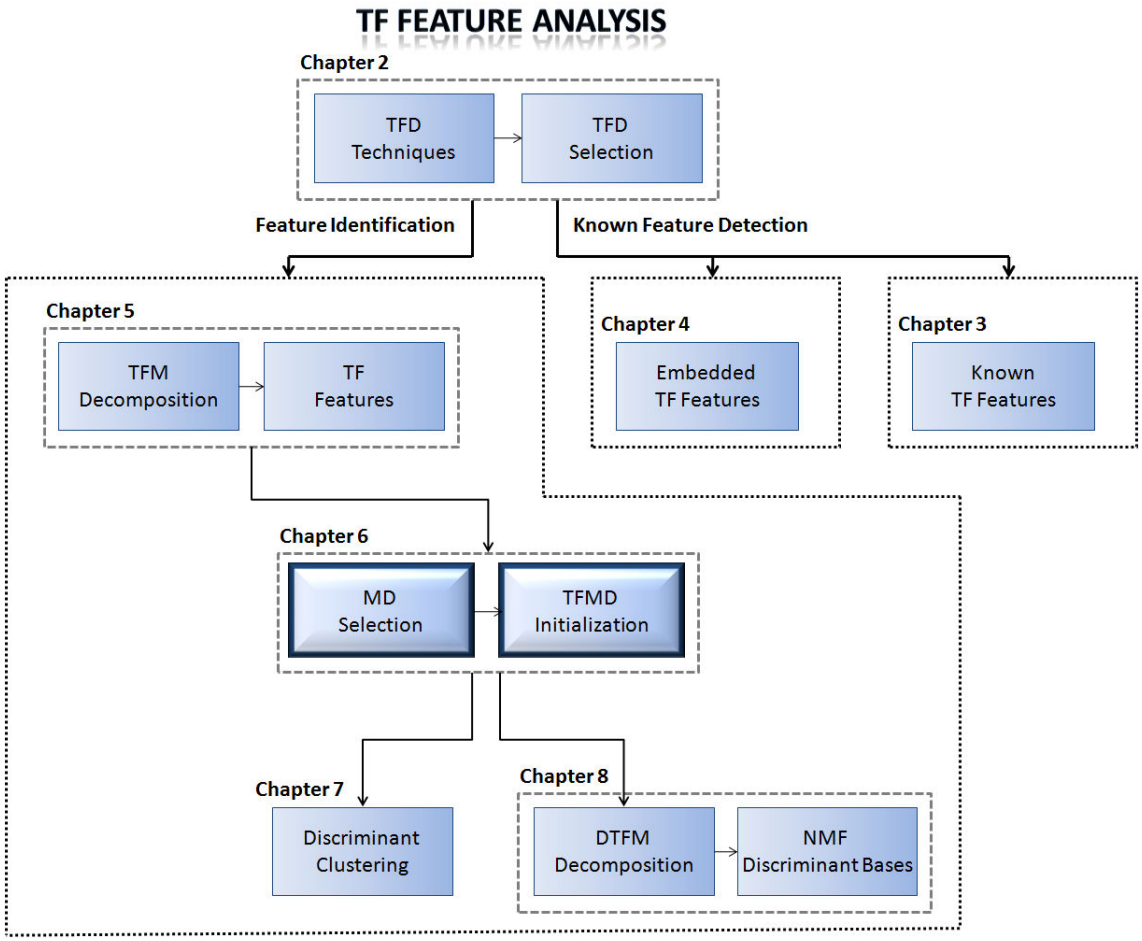


Figure 6.1: Chapter 6 - Selection of the matrix decomposition technique.

6.1 Motivation

THE previous chapter proposed TFM decomposition to adaptively characterize the TF structure in a given data. Although this approach successfully analyzes the non-stationarities in the data, its performance heavily depends on the quality of the matrix decomposition (MD) technique. Therefore, in this chapter, we investigate the well-known MD techniques as related to TFM quantification, selecting the most suitable technique to be used in the proposed pattern recognition system.

There are several MD techniques available including principal component analysis (PCA), independent component analysis (ICA), and non-negative matrix factorization (NMF). Performance comparison of ICA and PCA to NMF for feature extraction has been experimentally investigated for different applications; however, depending on the application and the data, different results have been reported. Some works demonstrated the advantages of ICA to NMF and PCA; for example, in [85], Lee et al. study PCA, NMF and ICA for feature extraction from multiple video frames. In this work, the authors show that ICA-based features result in better video representation than the PCA or NMF-based features. The author in [86] compares the techniques for occluded face recognition task, and shows that ICA method outperforms the other methods. Kim et al. [16] apply ICA, PCA and NMF to spectrogram of the signal to extract features to classify video sound track content. The result showed that NMF yields the lowest recognition rate for decomposition dimension of 23 or less ($r \leq 23$).

On the other hand, there are some studies that report the advantages of NMF to ICA or PCA. Chikhi et al. [87] compare the dimensionality reduction performance of the three techniques for web structure mining. They show that for basis smaller than 10, PCA, ICA and NMF show the same performances, but when increasing the number of bases, there is a significant degradation in the performance of ICA and PCA, while NMF remains stable. In [88], Cho et al. compare ICA and NMF for sound classification, and show that NMF based method outperforms the method based on ICA.

Considering the above contradictory reports on the performance of the techniques and lack of a comprehensive comparison among them as related to TF quantification, in this chapter, we focus

on investigating which MD results in TF features with a higher performance. Such a desirable MD should decompose a TFM into its significant components that correctly represent the TF structure. The contribution of this chapter is highlighted in the block diagram of Fig. 6.2. First, we introduce three well-known MD techniques. Next, we perform a fair comparison of the above three well-known techniques on a controllable dataset to select the most suitable technique for quantification of TF plane. Once we select the right MD, we integrate the TF analysis into the MD technique, and propose a novel seeding method for the MD optimization stage. Once we designed the matrix decomposition block the outcome is used in the pattern recognition system. We decompose the TFM ($\mathbf{V}_{M\times N}$) into its significant components ($\mathbf{W}_{M\times r}$ and $\mathbf{H}_{r\times N}$) as explained in Section 5.2, and then characterize them with the unique TF features proposed in Section 5.3. Finally, an application of the developed system for audio classification is demonstrated.

6.2 Matrix Decomposition Methods

This section reviews three well-known matrix decomposition techniques: PCA, ICA, and NMF.

6.2.1 Principal Component Analysis

The idea of PCA, also known as Karhunen-Loeve transformation, appeared about one hundred years ago [89]. The objective of PCA is to find a set of orthogonal components that minimizes the mean square error of the reconstructed data, and represent the original data with fewer components to reduce the dimensionality of the data. The PCA algorithm considers that the set of eigenvectors, $\mathbf{W}$, corresponding to the first r largest eigenvalues of the covariance matrix of the data, $\mathbf{V}$, are the principal components of the data set. These eigenvectors are called principle axes of the data, and the sub-space introduced by them is called principle component space. If $\mathbf{H}$ is the projection of the data on this space, the original data can be reconstructed as: $\hat{\mathbf{V}} \approx \mathbf{WH}$. The traditional PCA methods, which are based on least-square error, are not robust to outliers, such as noise. There are two ideas for robust PCA estimation including: calculating of eigenvalues and eigenvectors based on a robust estimate of the covariance matrix [90]; and computing the robust estimates of the eigenvalues and eigenvectors using projection pursuit [91]. The latter technique is used in the

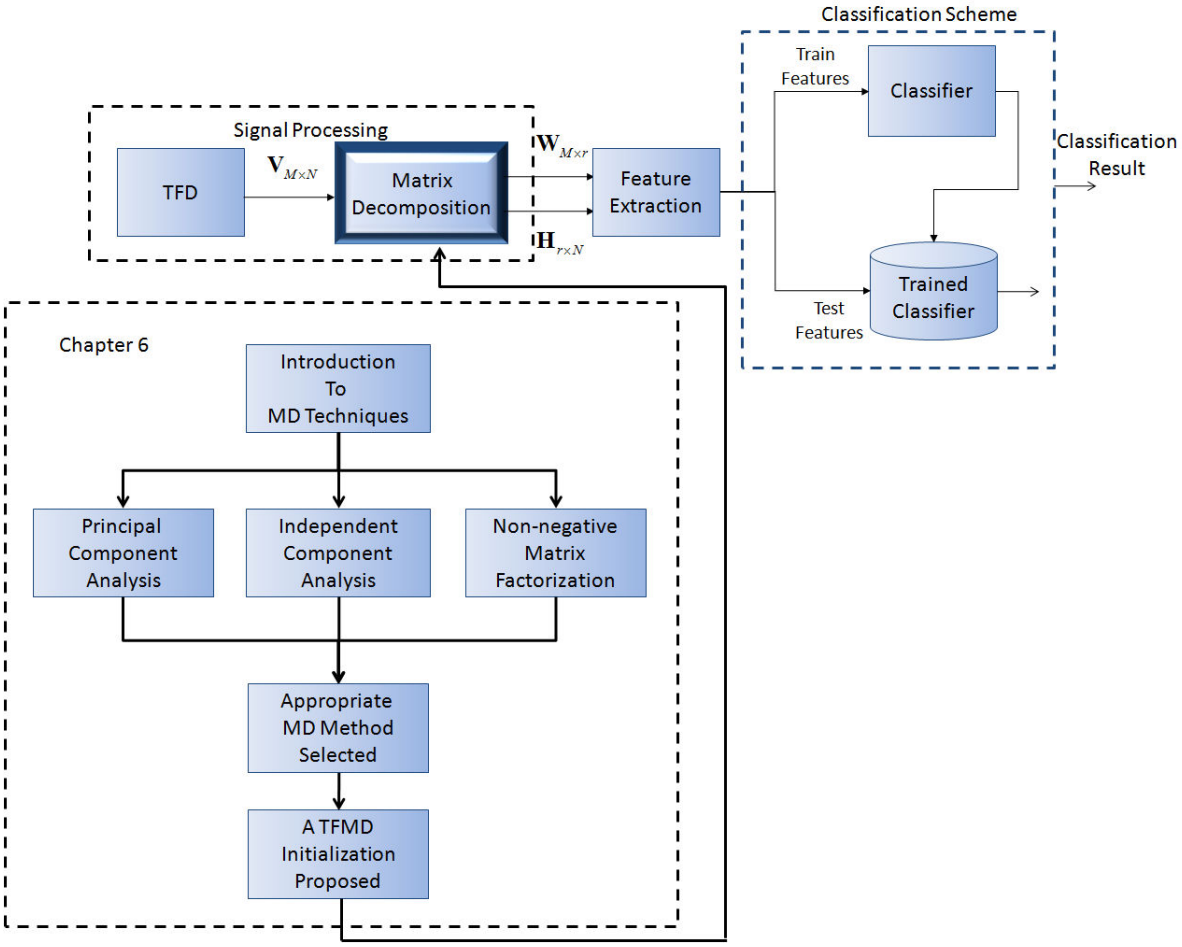


Figure 6.2: Chapter 6 - Matrix Decomposition.

present work, and it is denoted as Robust PCA (RPCA).

PCA reduces the dimension of the data while keeping the direction and the corresponding strength of major variations in the data, so it's being widely used for denoising, dimensionality reduction and data compression, feature extraction and classification [92, 93, 94, 95, 96, 97].

6.2.2 Independent Component Analysis

ICA is a statistical technique for decomposing a dataset into components that are as independent as possible. If r independent components $w_1 ... w_r$ compose r linear mixtures $v_1 ... v_n$ as $\mathbf{V} = \mathbf{HW}$, the goal of ICA is estimating $\mathbf{H}$, while our observation is only the random matrix $\mathbf{V}$. Once the matrix $\mathbf{H}$ is estimated, the independent component can be obtained as: $\mathbf{W} = \mathbf{H}^{-1}\mathbf{V}$. The fundamental idea of ICA estimation is as follow:

According to Central Limit Theorem, if $v = hw$ is a linear combination of independent components $w_1 ... w_r$, the vector h that maximizes the non Gaussianity of $h^T v$ corresponds to one of the independent components. Therefore, $h^T v$ could be said that is equal to one of the components, if its Gaussianity becomes minimum. This is true under the condition that all the independent components are non Gaussian; otherwise, the matrix $\mathbf{H}$ will not be identifiable. Hyvrinen and Oja proposed Fast ICA (FastICA) [98], which is one of the most efficient practical learning rules for ICA. FastICA is based on a fixed-point iteration scheme to find h_i with maximum non Gaussianity of $h_i^T v$ for component i. To estimate other independent components, the one-unit FastICA is run for each components. To prevent the algorithm from converging to the same component, the algorithm decorrelates the outputs $h_1^T v, ..., h_r^T v$ after every iteration. In this study, we utilize Fast ICA (FastICA) which is an efficient algorithm for ICA.

ICA has been broadly utilized for source separation and removal of artifacts and noise [99, 100, 101, 102].

6.2.3 Non-negative Matrix Factorization

NMF was performed in the middle of the 1990s under the name positive matrix factorization (PMF) [103]. In 1999, Lee and Seung [104] introduced some simple algorithms for the factorization, and

demonstrated the success of the technique on some classification applications. NMF decomposes a non-negative matrix, and constraints the matrix factors **W** and **H** to be non-negative; therefore, NMF tempts to extract part-based features. NMF algorithm starts with an initial estimate for **W** and **H**, and performs an iterative optimization to minimize a given cost function. In [105], Lee and Seung introduce two updating algorithms using the least square error and the Kullback-Leibler (KL) divergence as the cost functions:

$$\begin{aligned} &\text{Least square error:} \quad \mathbf{W} \leftarrow \mathbf{W}.\frac{\mathbf{V}\mathbf{H}^T}{\mathbf{W}\mathbf{H}\mathbf{H}^T}, \quad \mathbf{H} \leftarrow \mathbf{H}.\frac{\mathbf{W}^T\mathbf{V}}{\mathbf{W}^T\mathbf{W}\mathbf{H}}, \\ &\text{KL divergence:} \quad \mathbf{W} \leftarrow \mathbf{W}.\frac{\frac{\mathbf{V}}{\mathbf{W}\mathbf{H}}\mathbf{H}^T}{\mathbf{1}.\mathbf{H}}, \quad \mathbf{H} \leftarrow \mathbf{H}.\frac{\mathbf{W}^T\frac{\mathbf{V}}{\mathbf{W}\mathbf{H}}}{\mathbf{W}.\mathbf{1}}, \end{aligned} \tag{6.1}$$

In these equations, $\langle\ .\ \rangle$ and $\frac{\langle\rangle}{\langle\rangle}$ are term by term multiplication and division of two matrices, and $\mathbf{1}$ is a matrix of ones. The least square error approach is a standard bound-constrained optimization problem. However, KL divergence formula is not a bound-constrained problem, which requires the objective function to be well-defined at any point of the bounded region [106]. The log function is not well-defined if any elements in matrix **V** or **WH** is zero. Hence, we do not consider KL divergence formulation in this study.

NMF Optimization

Various alternative minimization strategies have been proposed [107]. In this work, we use a projected gradient bound-constrained optimization method which is proposed by Lin [106]. The optimization method is performed on function $f = \mathbf{V} - \mathbf{WH}$, and is consisted of three steps:

1) *Updating the Matrix **W***: In this stage, the optimization of $f_{\mathbf{H}}(\mathbf{W})$ is solved respect to **W**, where $f_{\mathbf{H}}(\mathbf{W})$ is the function $f = \mathbf{V} - \mathbf{WH}$, in which matrix **H** is assumed to be constant. In every iteration, matrix **W** is updated as below:

$$\mathbf{W}^{t+1} = \max\left\{\left(\mathbf{W}^t - \alpha^t \nabla f_{\mathbf{H}}(\mathbf{W}^t)\right), 0\right\} \tag{6.2}$$

where t is the iteration order, $\nabla f_{\mathbf{H}}(\mathbf{W})$ is the projected gradient of the function f, while **H** is constant and α^t is the step size to update the matrix. The step size is found as $\alpha^t = \beta^{K_t}$. Where $\beta^1, \beta^2, \beta^3, \ldots$ are the possible step sizes, and K_t is the first non-negative integer for which:

$$f(\mathbf{W}^{t+1}) - f(\mathbf{W}^t) \leq \sigma\left\langle \nabla f_{\mathbf{H}}(\mathbf{W}^t), \mathbf{W}^{t+1} - \mathbf{W}^t \right\rangle \tag{6.3}$$

where the operator $\langle ., . \rangle$ is the inner product between two matrices as defined below:

$$\langle A, B \rangle = \sum_i \sum_j a_{ij} b_{ij} \tag{6.4}$$

In [106], values of σ and β are suggested to be 0.01 and 0.1, respectively. Once the step size, α^t, is found, the stationarity condition of function $f_{\mathbf{H}}(\mathbf{W})$ at the updated matrix is checked as below:

$$\left\| \nabla^P f_{\mathbf{H}}(\mathbf{W}^{t+1}) \right\| \leq \epsilon \left\| \nabla f_{\mathbf{H}}(\mathbf{W}^1) \right\| \tag{6.5}$$

where $\left\| \nabla f_{\mathbf{H}}(\mathbf{W}^1) \right\|$ is the the projected gradient of the function $f_{\mathbf{H}}(\mathbf{W})$ at first iteration ($t = 1$), ϵ is a very small tolerance, and $\nabla^P f_{\mathbf{H}}(\mathbf{W})$ is the projected gradient defined as:

$$\nabla^P f_{\mathbf{H}}(\mathbf{W}) = \begin{cases} \nabla f_{\mathbf{H}}(\mathbf{W}), & w_{mr} > 0, \\ \min\left(0, \nabla f_{\mathbf{H}}(\mathbf{W})\right), & w_{mr} = 0, \end{cases} \tag{6.6}$$

If the stationary condition is met, the procedure stops, if not, the optimization is repeated until the point $\mathbf{W}^{t+1}$ becomes a stationary point of $f_{\mathbf{H}}$.

2) *Updating the Matrix* ***H***: This stage solves the optimization problem respect to **H** assuming **W** is constant. A similar procedure to what we did in stage 1 is repeated in here. The only difference is that in the above stage, **H** is constant, but in here **W** is constant.

3) *The convergence test*: Once the above sub-optimum problems are solved, we check for the stationarity of the **W** and **H** solutions together:

$$\begin{aligned} &\left\| \nabla f_{\mathbf{H}}(\mathbf{W}^t) \right\| + \left\| \nabla f_{\mathbf{W}}(\mathbf{H}^t) \right\| \\ &\leq \epsilon \left(\left\| \nabla f_{\mathbf{H}}(\mathbf{W}^1) \right\| + \left\| \nabla f_{\mathbf{W}}(\mathbf{H}^1) \right\| \right) \end{aligned} \tag{6.7}$$

The optimization is complete if the global convergence rule (Eqn. 6.7) is satisfied; otherwise, the steps 1 and 2 are iteratively repeated until the optimization is complete.

The gradient-based NMF is computationally competitive and offers better convergence properties than the standard approach, and it is, therefore, used in the present study.

6.3 Selection of Matrix Decomposition Technique

In this section, we select the most appropriate MD technique for TF quantification application. To make this happen, we compare the performance of the three well-known MD techniques as

explained in the previous section. Two experiments are performed; in the first experiment, we explore which MD method provides the most accurate decomposition at the decomposing stage. In the second experiment, the TF features obtained using different MD methods are compared as related to their TF localization performance. Based on the result, we select the suitable MD technique for the pattern recognition applications.

6.3.1 Experiment 1: TFM Decomposition

In order to measure the decomposition accuracy of each MD technique, each technique is used to decompose the TFM of a synthetic signal. The synthetic signal is generated using Eqn. 8.16. Next, the extracted base and the coefficient matrices are used to reconstruct the TFM, denoted by $\mathbf{V}_{rec}$. Then, using the reconstructed matrix and the pre-known instantaneous frequency (IF) information of the signal, we propose a figure of merit to measure the decomposition performance of each MD technique. The moment-based figure of merit is selected in this study as it calculates the reconstruction accuracy according to both the structure and the energy of the samples.

Figure of Merit

The synthetic signal generated using Eqn. 5.1 is composed of seven components, and each component j ($j = 1, ..., 7$) is characterized with 4 parameters: beginning time, t_{j1}, ending time, t_{j2}, and the line parameters (a_j and b_j) which represent the IF of each component:

$$IF_j(n) = a_j n + b_j, \tag{6.8}$$

Fig. 6.3 demonstrates the IF parameters for a linearly frequency modulated component. For each component, we define the first order moment of the reconstructed TF matrix around its IF as the localization of that component:

$$\text{Lcz}_j = \sum_{n=t_{j1}}^{t_{j2}} \sum_{m=IF_j(n)+\Delta}^{IF_j(n)-\Delta} \left(|\mathbf{V}_{rec}(m,n)| \times |m - IF_i(n)|\right), \tag{6.9}$$

where 2Δ (shown in Figure 6.3) is the frequency interval around the IF in which we calculate the localization. The percentage localization of component j is calculated as:

$$\text{Localization}_j(\%) = 100 - \left(\frac{|\text{Lcz}_j - \text{Lcz}_{O-j}|}{\text{Lcz}_{O-j}} \times 100 \right), \tag{6.10}$$

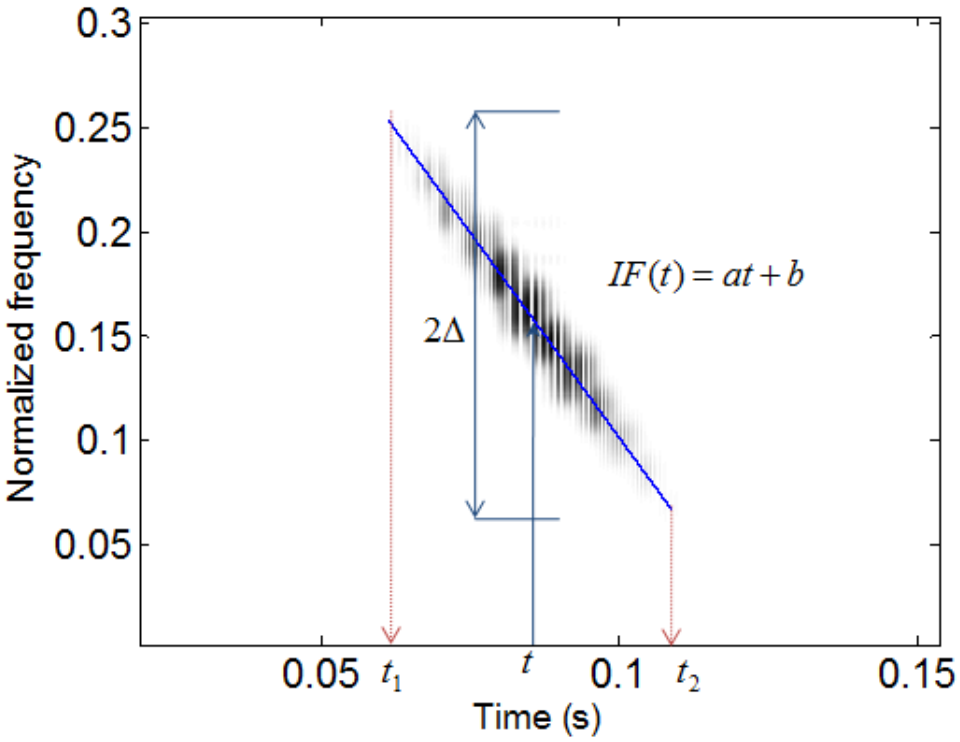


Figure 6.3: Each component in the synthetic signal is defined with 4 parameters: start time t_1, ending time t_2 and the parameters of the IF line a and b.

where, Lcz_j and Lcz_{O-j} are the localization values calculated using Eqn. 6.9 from the reconstructed and the original TF component j, respectively, and Localization_j is the quantified localization of that component.

When MD provides an accurate decomposition of the TFM, $|\text{Lcz}_j - \text{Lcz}_{O-j}|$ will be a very small value and as a result Localization_j in Eqn. 6.10 will be very close to 100%, but if MD does not estimate an accurate decomposition, Localization_j will be small value.

Results

The proposed figure of merit is applied to evaluate the decomposition performances of the three MD techniques as described below:

1. Ten synthetic signals with varying characteristics are generated.
2. Adaptive TFD of each signal is constructed.
3. Robust PCA (RPCA), Fast ICA (FICA) and NMF decompose the TF matrix to **W** and **H**.
4. **W** and **H** are used to reconstruct the TF matrix: $\mathbf{V}_{rec} = \mathbf{WH}$
5. Localization of each component (Localization_i) is calculated using Eqn. 6.9

The average localization percentage of each technique is plotted in Fig. 6.4. In this figure, the localization percentage of each component type is calculated separately; the first three bars belong to the frequency localized components (components 1, 3 and 4 in Fig. 5.10 (b)), respectively. The next three bars represent components number 2, 5 and 6, and the last three bars belong to component 7.

6.3.2 Experiment 2: TF Features

We compare the representation and localization of the extracted TF features using PCA, ICA and NMF as the MD techniques. The synthetic signal shown in Fig. 6.5(a) is used in the example. This signal is constructed of three components: a frequency localized component, a transient and a frequency linearly modulated component. The TFM features of the signal are extracted as explained

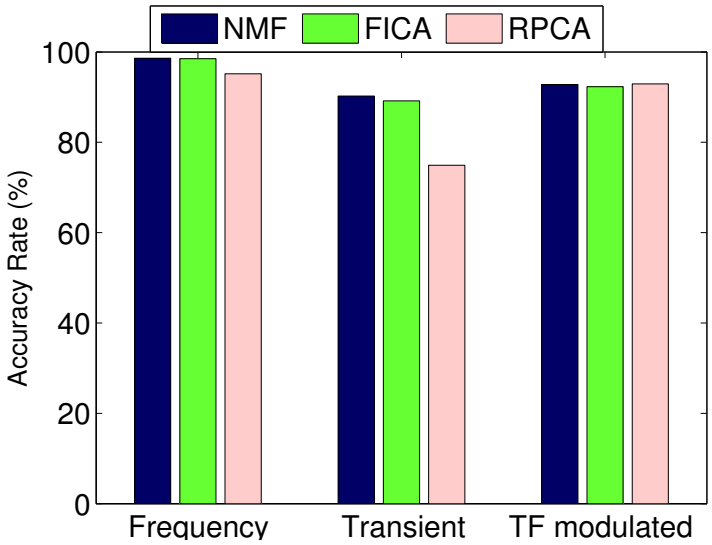


Figure 6.4: Localization performance of NMF, ICA and PCA are compared for frequency localized, transient and frequency modulated components.

in the following steps:

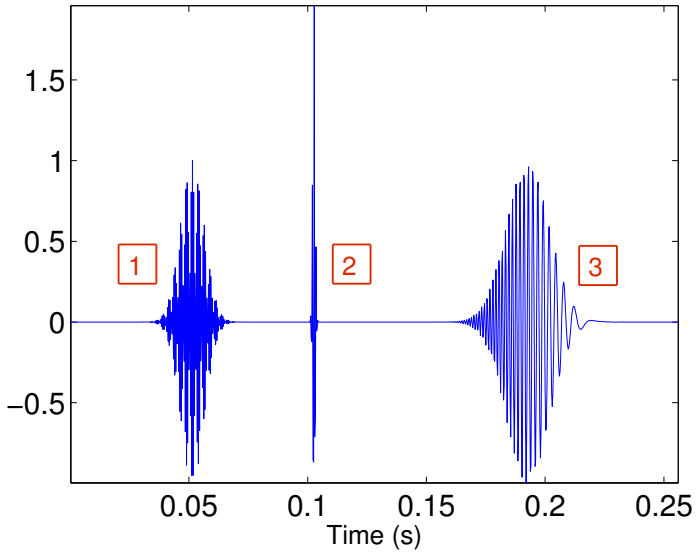
1. The TFM, $\mathbf{V}_{M\times N}$,is constructed using adaptive TFD technique (As explained in Section 2.1), and is illustrated in Fig. 6.5(b).

2. TFM decomposition is performed as the following equation: $\mathbf{V}_{M\times N} = \mathbf{W}_{M\times r}\mathbf{H}_{r\times N}$.

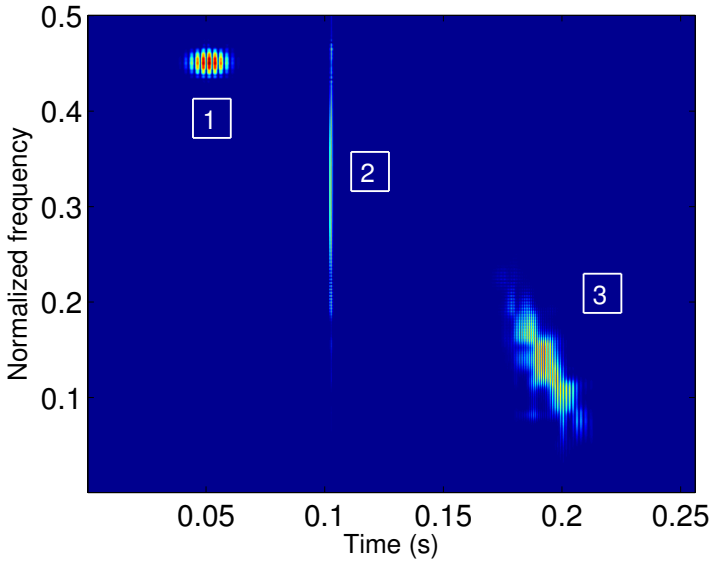
3. As it was mentioned in Section 5.3, spectral and temporal moments should be extracted from non-negative vectors. PCA and ICA techniques do not guarantee the non-negativity constraints, therefore, instead of using $\mathbf{W}$ and $\mathbf{H}$ matrices directly, their squared values are used; denoted by $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{H}}$, respectively.

4. Finally, the first and second order of spectral and temporal moments are extracted as shown in the following:

$$\begin{aligned} F_j &= \left(\hat{e}_j, \bar{t}_j, \bar{f}_j, \hat{t}_j, \hat{f}_j\right) \\ &= \left(10\log\left(e_j/e_{max}\right), \langle t\rangle_j, \langle f\rangle_j, \sqrt{\langle t^2\rangle_j - \bar{t}_j^2}, \sqrt{\langle f^2\rangle_j - \bar{f}_j^2}\right), \end{aligned} \tag{6.11}$$



(a) Synthetic signal in time domain.



(b) TFD of the synthetic signal.

Figure 6.5: A synthetic signal with three components: from left to right the components are frequency localized, transient and frequency modulated. TFM is constructed using adaptive TFD with Gabor atoms and 100 iterations and MCE of 5 iterations.

where,

$$\left\langle t^{(p)} \right\rangle_j = \sum_{n=0}^{N} n^{(p)} h_j(n),$$
$$\left\langle f^{(q)} \right\rangle_j = \sum_{m=0}^{M} m^{(q)} w_j(m), \tag{6.12}$$

and e_j is the average energy of the signal in the rectangle feature j,

$$e_j = \text{mean} \left[\sum_{n=\bar{t_j}-\hat{t_j}}^{\bar{t_j}+\hat{t_j}} \sum_{m=\bar{f_j}-\hat{f_j}}^{\bar{f_j}+\hat{f_j}} \mathbf{V}_{rec}(m,n) \right] \tag{6.13}$$

and e_{max} is the feature vector with highest energy.

Results

Tables 6.1,6.2 and 6.3 present the derived features utilizing PCA, ICA and NMF respectively. The feature vectors, F_1 to F_7, are sorted from the highest energy to the lowest energy. The numbers which are displayed in the last row of each feature vector indicate the component which that feature vector represents; for example, the first feature (F_1) in Table 6.1 is associated with the third component in Fig. 6.5 (The components in Fig. 6.5 are numbered 1, 2 and 3 from left to right).

Table 6.1: TF Features are extracted using PCA as the Matrix Decomposition tool.

Parameters	**Features**						
	F_1	F_2	F_3	F_4	F_5	F_6	F_7
$\hat{e}$	0	-0.5543	-2.9342	-3.6079	-17.2361	-17.6982	-19.5239
$\bar{f}$	0.1230	0.3209	0.1206	0.1522	0.1624	0.2805	0.3143
$\bar{t}$	0.1943	0.1025	0.1974	0.1894	0.1897	0.0893	0.0787
$\hat{f}$	0.0286	0.0530	0.0528	0.0335	0.0673	0.1585	0.1582
$\hat{t}$	0.0064	0.0010	0.0111	0.0077	0.0100	0.0628	0.0559
j	3	2	3	3	3	-	2

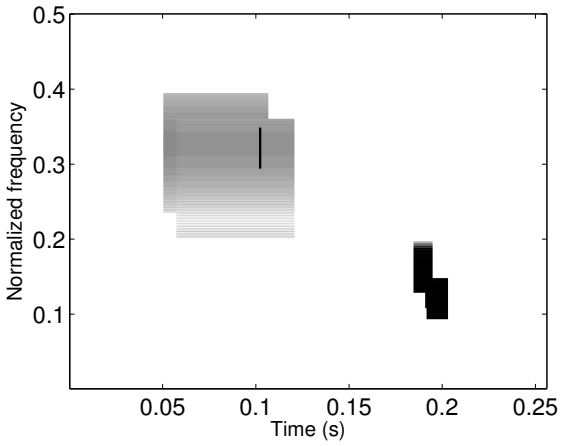
In addition, Figs. 6.6(a), 6.6(b) and 6.6(c) provide a comprehensible demonstration of the extracted feature vectors. The images in Figs. 6 (a), (b) and (c) follow from the feature set which is extracted using Eqn. (13), and are shown in Tables I, II and III. Each feature vector is associated with a rectangular region centered at $\bar{t}$ and $\bar{f}$, and width of $\hat{t}$ and $\hat{f}$ in time and frequency, respectively. The feature set consists of only 35 values, which is a tremendous reduction in data size, yet it still retains considerable detail covering the time-frequency energy distribution.

Table 6.2: TF Features are extracted using ICA as the Matrix Decomposition tool.

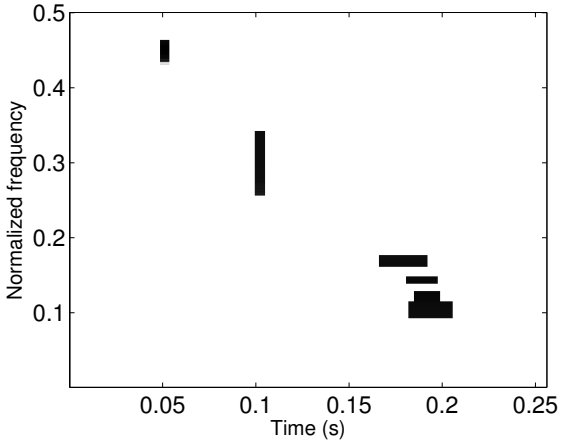
Parameters	**Features**						
	F_1	F_2	F_3	F_4	F_5	F_6	F_7
$\hat{e}$	0	-3.1425	-3.3637	-6.1000	-8.7705	-8.8756	-43.1515
$\bar{f}$	0.4493	0.1218	0.1443	0.1038	0.2993	0.1700	0.0852
$\bar{t}$	0.0515	0.1929	0.1903	0.1957	0.1024	0.1812	0.1484
$\hat{f}$	0.0146	0.0116	0.0170	0.0194	0.0852	0.0215	0.0359
$\hat{t}$	0.0067	0.0132	0.0118	0.0197	0.0051	0.0190	0.0512
j	1	3	3	3	2	3	-

Table 6.3: TF Features are extracted using NMF as the Matrix Decomposition tool.

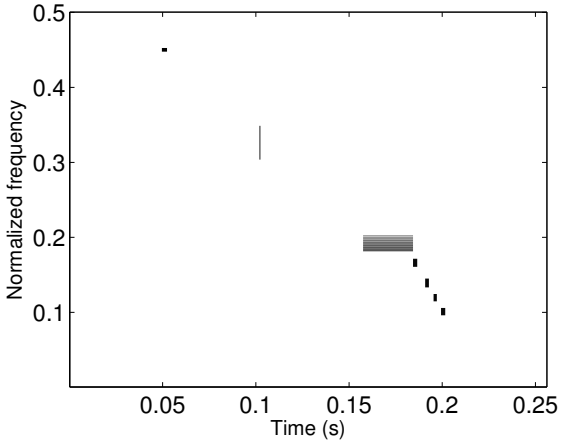
Parameters	**Features**						
	F_1	F_2	F_3	F_4	F_5	F_6	F_7
$\hat{e}$	0	-1.5642	-3.8209	-4.2875	-6.4465	-7.4118	-8.0806
f	0.4488	0.4511	0.1193	0.1397	0.3257	0.1021	0.1678
$\bar{t}$	0.0512	0.0513	0.1964	0.1918	0.1025	0.2006	0.1855
$\hat{f}$	0.0020	0.0049	0.0082	0.0105	0.0450	0.0081	0.0106
$\hat{t}$	0.0025	0.0031	0.0014	0.0019	0.0008	0.0019	0.0021
j	1	1	3	3	2	3	3



(a) PCA-based TF features.



(b) ICA-based TF features.



(c) NMF-based TF features.

Figure 6.6: The extracted features are plotted in TF plane. Each rectangle represents one feature vector.

6.3.3 MD Selection

From the tables and the figures presented above, some interesting properties are concluded as explained in the following:

- The first observation is that NMF features are highly localized in TF plane, while PCA and ICA-based features are more spread. The reason is that since the bases and coefficients of ICA and PCA are not necessarily non-negative, we had to extract the moment features from the squared-value base and coefficient vectors. In other words, we are extracting the features from

$$\hat{\mathbf{V}}_{rec}(m,n) \approx \sum_{j=1}^{r} |w_j(m)|\,|h_j(n)|\,, \tag{6.14}$$

  rather than $\mathbf{V}_{rec} \approx \mathbf{WH}$. Therefore, the negative elements of $\mathbf{W}$ and $\mathbf{H}$ matrices cause artifacts in the reconstructed TF matrix. The 6th feature vector in Table 6.1, and the 7th one in Table 6.2 are artifact feature vectors caused by the negativity of the ICA and PCA-based decompositions as the test signal does not have any energy at these locations.

- Another observation is the limitation of PCA in localization of component 1. The last row of Table 6.1 and Figure 6.6(a) indicates that none of the PCA features represent component 1. PCA decomposes the TF images into orthogonal bases, and successfully reconstructs the TF matrix; however, because of the presence of negative elements in the decomposed matrices, the moment-based features are not able to localize the first component.

- Additionally, NMF provides the most efficient data reduction compared to ICA and PCA. In this example, we extracted 7 signatures from the TFD; however, the last three features contain only 1.6% of the total feature energy, and only the first 4 features obtained from the PCA method are useful features. The same conclusion applies to ICA as the last feature in Table 6.2 is too small to be effective in TF classification. It can be observed from Figure 6.6 that only NMF results in 7 representative TF features.

Table 6.4 summarizes the properties of the three well-known MD techniques as related to characterization of TFD.

Table 6.4: Desirable MD Properties for TF Quantification. The more are the number of the stars at each property indicates that the method is more desirable with respect to that specific property.

Property	Localized Reconstruction	TF Feature Localization	Transient Characterization	Efficient Representation
PCA	*	*	*	*
ICA	**	**	***	**
NMF	***	***	***	***

Based on the above and as our goal is to extract the TF features that successfully characterize the TF structure of a signal, we select NMF as the MD in our developed system. However, a common shortcoming of all the NMF optimization approaches is initialization of **W** and **H** matrices. The NMF decomposition algorithms start with random initialization for **W** and **H**, and modify these two matrices iteratively until the cost function is minimized. However, due to the non-convexity of the cost function in both **W** and **H**, depending on the initial matrices, at each optimization, a different local minima of the cost function may be achieved. As a result, each time we run the algorithm, NMF might result in a different decomposition output. It was theoretically shown in [108] that under some conditions, the NMF decomposition will be unique, but it is also shown that these conditions are not generally satisfied on the case of a real world data. One solution to this shortcoming is appropriate seeding of **W** and **H** matrices. Various efforts have been performed to find alternate seeding approaches in order to influence the NMF convergence to a desired solution. In [109], spherical k-means clustering is employed to initialize **W**, and in [110], SVD is used to seed NMF matrices. To address the initialization problem of the NMF algorithm, we propose a novel seeding method as is explained in the following section.

6.4 Initialization of TFM Decomposition

Our motivation in the proposed NMF seeding method is to use our knowledge about the TF structure of a signal to find suitable initialization values for **W** and **H** matrices. As shown in Eqn. 2.12, the MP-TFD of a signal ($x(t)$) is constructed as the addition of the WVD of the first I Gabor functions. Using the elementary properties of WVD (page 107 in [37]), the WVD of a Gabor function

G_{γ_i} in Eqn. 2.10 can be written as follows:

$$\mathbf{WVG}_{\gamma_i}(t,f) = \mathbf{WVg}(\frac{t-p_i}{s_i}, s_i(f-f_i)), \tag{6.15}$$

where $\mathbf{WVg}(t,f)$ is the WVD of the Gaussian function $g(t)$ in Eqn. 2.10. As shown in Example 4.18 in [37], in case of Gabor atoms, the WVD of a Gaussian atom is a two-dimensional Gaussian which can be written as follows:

$$\mathbf{WVg}(t,f) = \hat{g}^2(f)g(t)^2, \tag{6.16}$$

where, $\hat{g}(f)$ is the Fourier transform of $g(t)$. In page 28 of [37], it is shown that the Fourier transform of a Gaussian function is also a Gaussian.

From Eqn. 6.16, matrix **WVg** can be shown as follows:

$$\mathbf{WVg} = \begin{bmatrix} \hat{g}^2(1) \\ \hat{g}^2(2) \\ \vdots \\ \hat{g}^2(M) \end{bmatrix} \begin{bmatrix} g^2(1)g^2(2)\cdots g^2(N) \end{bmatrix} \tag{6.17}$$

In the above equation, **WVg** is an $M \times N$ matrix, where N and M are the number of samples in $x(t)$ and the frequency resolution, respectively.

We combine Eqns. 6.15 and 6.17 in a way that the MP-TFD in Eqn. 2.12 can be written in a matrix format as follows:

$$\mathbf{V}(t,f) = \sum_{i=1}^{I} |a_{\gamma_i}|^2 \begin{bmatrix} \hat{g}^2(s_i(1-f_i)) \\ \hat{g}^2(s_i(2-f_i)) \\ \vdots \\ \hat{g}^2(s_i(M-f_i)) \end{bmatrix} \times \tag{6.18}$$
$$\left[g^2(\frac{1-p_i}{s_i}) g^2(\frac{2-p_i}{s_i}) \cdots g^2(\frac{N-p_i}{s_i}) \right],$$

Comparing the matrix display of MP-TFD (Eqn. 6.19) with the TFM decomposition shown in Eqn. 2.11, we can see that the Gabor-based MP-TFD provides a decomposition of the TFD with decomposition order of I (r in Eqn. 2.11 is equal to I). Previously, we mentioned that in order to decompose all the coherent structures in a signal, the number of iterations has to be very large; for example for a 3s audio signal, 1000 iterations are needed. Therefore, the TFM decomposition

that is performed through MP-TFD is a very redundant decomposition (r is very large). According to the nature of matching pursuit algorithm, at every iteration some portion of the signal energy is modeled with an optimal TF resolution in the TF plane, and the first few iteration contains the best correlated TF functions selected from the Gabor dictionary. Therefore, we assume that the first few atoms in MP-TFD decomposition provide significant information about the TF structure of the signal, and we propose to use the first r Gabor atoms in Eqn. 6.19 to initialize the NMF algorithm as follows:

$$w_i^{\textit{Init}} = |a_{\gamma_i}| \begin{bmatrix} \hat{g}^2(s_i(1-f_i)) \\ \hat{g}^2(s_i(2-f_i)) \\ \vdots \\ \hat{g}^2(s_i(M-f_i)) \end{bmatrix}, \tag{6.19}$$

$$h_i^{\textit{Init}} = |a_{\gamma_i}| \left[g^2(\tfrac{1-p_i}{s_i}) g^2(\tfrac{2-p_i}{s_i}) \cdots g^2(\tfrac{N-p_i}{s_i}) \right],$$
$$\text{for } i = 1, ..., r$$

The proposed seeding method is based on the TF structure of the signal, and it is therefore expected to result in a faster convergence. Additionally, unlike the random initialization that results in a different local minima of the cost function every time we repeat the algorithm, the proposed initialization technique achieves one unique decomposition output. In order to evaluate the proposed seeding method, we used the MP-TFD based and random initialization methods to decompose an 80 ms of a speech signal. Fig. 6.7 depicts the mean squared error (MSE) between the original and the reconstructed TFM using each initialization methods. As shown in this figure, the proposed MP-TFD based initialization method starts with a larger MSE compared to the random initialization, but it converges a lot faster than the random seeding. After 8 iterations, NMF with MP-TFD initialization reaches to MSE of 0.01, while the random initialization method requires 18 iterations to achieve the same MSE. This experiment showed that the proposed initialization technique speeds up the convergence of NMF to the desired decomposed matrices.

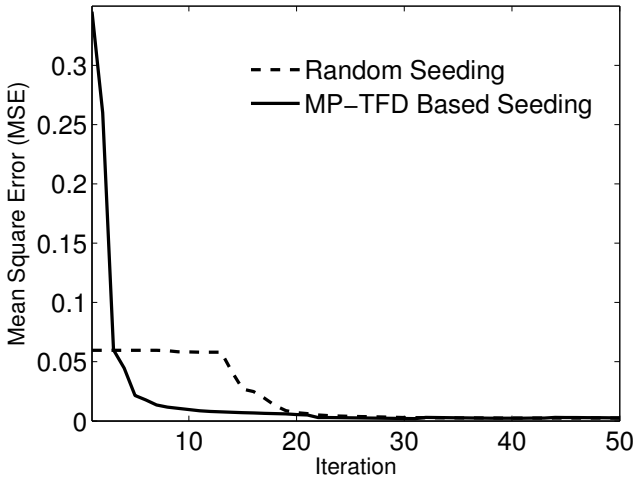


Figure 6.7: NMF Convergence is compared for MP and random-based seedings. Using the proposed initialization technique, NMF reaches to a MSE of 0.01 after 8 iterations, while with the random initialization, it takes 18 iterations to achieve the same MSE.

6.5 Selection of the number of components.

The number of TF components influences the quality of the proposed TF quantification technique. Wrong selection of this number might result in over or under analysis of the TFD. However, selection of the right number is strongly dependent on the length of the input TFM and its TF resolution (i.e., the dimension of the TFM: M and N). The other important factor in selection of the efficient decomposition number is the nature of the application in hand. Highly non-stationary signals require more decompositions compared to signals with less TF varying structures. In this dissertation, we experimentally choose these variables depending on the characteristics of the dataset in hand.

6.6 Experiment1: Environmental Audio Classification

Environmental audio classification is selected as one of the applications of our developed TFM feature extraction. Audio feature extraction and classification are important tools for audio signal analysis for many applications, such as multimedia indexing and retrieval, and auditory scene analysis. In this chapter, we employ the newly proposed approach to audio feature extraction in an

attempt to achieve high classification accuracy of environmental audio signals.

6.6.1 Background

The general methodology of audio classification involves extracting discriminatory features from the audio data and feeding them to a pattern classifier. The better and more effective features are extracted from audio signals, the more accurate results will be achieved in the audio classification technique. Over the last several years, different audio feature extraction techniques have been introduced. In general, all the feature extraction methods utilize one of the following three signal representation domains: temporal domain, spectral or joint time-frequency (TF) distribution. *Temporal domain features*, such as, signal energy, pitch, zero crossing rate [111, 112] and Entropy modulation [113] have been used for audio classification. Although temporal features have been traditionally applied for feature extraction applications, they are not enough to represent the non-stationary characteristics of audio signals.

Examples of *Spectral features* include 4 Hz modulation energy, percentage of low-energy frames, spectral rolloff point, spectral centroid, mean frequency, cepstral coefficients [114, 115], and high and low frequency slopes [116]. Additionally, since audio signals show different structural behavior from one frame to the next frame, several features are introduced to characterize the spectral difference between the neighboring frames. For example, spectrum flux (SF) [114] is defined as the average variation value of spectrum between the two adjacent frames. Although spectral features are useful in audio classification, they do not provide any information about the temporal evolution or the localization of the extracted features over a frame. Therefore, spectral features are not enough to represent environmental audio signals, specially the artificially created sounds such as music that indicate a highly non-stationary nature. There have been some attempts to derive joint *TF features* [117, 118, 119, 120, 121] from audio signals. TF features are effective for revealing non-stationary aspects of signals such as trends, discontinuities and repeating patterns where the two previous approaches fail or are not as effective.

Various classifiers have been utilized for audio classification. Audio content analysis at Microsoft research commonly uses Gaussian mixture models (GMM) [122], k-nearest neighborhood

(K-NN) [123] and support vector machine (SVM) [124] for audio classification. Other popular classifiers for audio classification include linear discriminant analysis (LDA) [125], hidden Markov models (HMM) [126] and artificial neural networks (ANN) [127]. There are some works that focus attention on developing new classifiers, or comparing existing classifiers for audio classification applications. For instance, in [128], Buchler et. al. compare simple classifiers (e.g., rule-based and minimum-distance classifiers) with complex approaches (e.g., Bayes classifier, neural network and hidden Markov model). While these studies are beneficial, the aim of the present study focuses on improving the quality of the extracted features. Therefore, in this section, we avoid complex classifiers and apply LDA as a simple linear classifier to evaluate our proposed features.

The lack of a common dataset does not allow a fair comparison of different audio classification methodologies. Some literatures report an impressive accuracy rate, but they use only a small number of classes and/or a small dataset in their evaluations. In this work we use an environmental audio dataset that was developed and compiled in our signal analysis research (SAR) group at Ryerson University [125]. The dataset is designed to have 10 different classes containing human speech, nature sounds, artificial sounds and three different types of musical instruments. More detailed characteristics of this dataset are explained in Section 6.6.3.

6.6.2 Methodology

Fig. 6.8 depicts the schematic of our proposed TF feature extraction methodology.

Feature Extraction

As it is depicted in this figure, once the TF matrix is decomposed into base and coefficient vectors, we extract 20 features from each base vector and its corresponding coefficient vector. Mel-frequency cepstral coefficient (MFCC) features are known to perform well with audio signals. Therefore, 13 of these features are the first 13 MFCC of each base vector. The next six features are $\mathrm{S}_h, \mathrm{S}_w, \mathrm{D}_h, \mathrm{D}_w, \mathrm{MO}_h$, and MO_w, which are extracted from base and coefficient vectors in a way that the obtained features represent discontinuities and transients in time and frequency, and the last feature MP is calculated from MP decomposition.

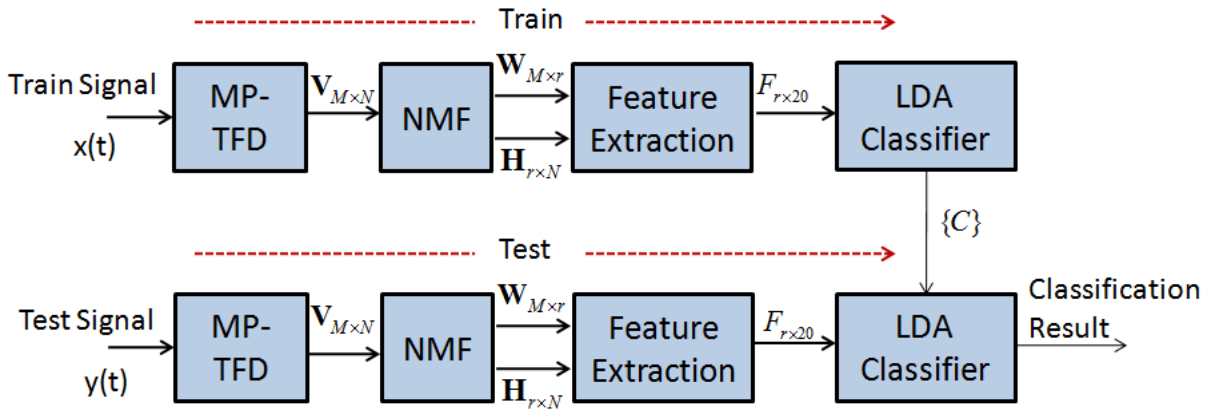


Figure 6.8: The block diagram of the proposed feature extraction methodology.

Classifier

Linear Discriminant Analysis (LDA) is applied as the classifier shown in Fig. 6.8. In discriminant analysis, the feature vector derived as explained above were transformed into canonical discriminant functions such as

$$T = f_1 b_1 + f_2 b_2 + ... + f_{20} b_{20} + a \tag{6.20}$$

where $\{f\}$ is the set of TF features, and $\{b\}$ and a are the coefficients and constant respectively. Using discriminant scores and the prior probability values of each group, the posterior probabilities of each sample occurring in each of the groups are computed. The sample is then assigned to the group with the highest posterior probability.

In simple terms in the feature space with dimensions equal to the number of features, linear planes are introduced which will divide the dataset into different groups. The covariances and probabilities are used to confine the area in the space where each class of signals occur. Once this area is defined, statistical distances are calculated between the centroid of each class, and linear planes are introduced in a optimal way to segregate the classes. Fig. 6.9 displays an example of LDA classifier with 2-D features and two-class classification indicated with H_0 and H_1. The classifier $\{b_2, b_1, a\}$ (i.e., the dashed line shown in the feature plane) divided the feature space into two spaces. For any given feature vector, the value of T is calculated as $T = f_2 b_2 + f_1 b_1 + a$. The

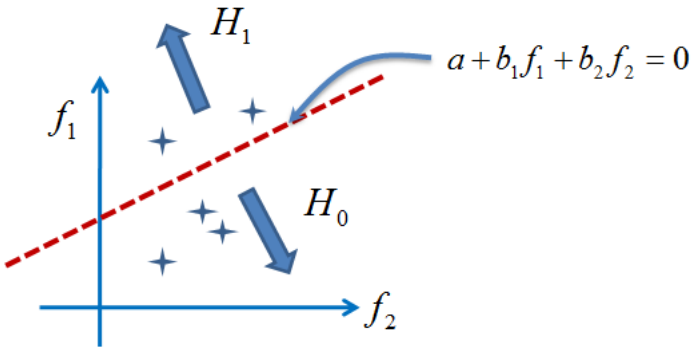


Figure 6.9: The schematic of LDA for 2 group classifier (H_0 and H_1) and 2-D feature space (f_1 and f_2).

feature belongs to group H_1 if the feature point is located above the classifier line, i.e., $T > 0$, or it belongs to H_0 if $T < 0$ and the feature point is under the classifier line.

6.6.3 Audio Database

The number of classes used in literature varies from study to study; for example, in [129], the authors use two classes (i.e., speech and music) while audio content analysis at Microsoft research [123] uses four audio classes (i.e., speech, music, environment sound, and silence). Freeman et al [127] use four classes of speech (i.e., babble, traffic noise, typing, and white noise). The authors in [120] use 14 different environmental scenes (i.e., inside restaurants, playground, street traffic, train passing, inside moving vehicles, inside casinos, street with police car siren, street with ambulance siren, nature-daytime, nature-nighttime, ocean waves, running water, raining, and thundering). In general, there are two types of audio classifications. In the first type of audio classification, the goal is classifying audio signals that belong to different families; for example, music and human sound belong to different families. In the second type, the purpose is classification of the signals that belong to one general family; for instance music classification belongs to the second type of audio classification.

We designed a data set in our signal analysis research (SAR) group at Ryerson University [125] such that it consists of both classification problems. This database consists of 192 audio signals of 5 s duration each with a sampling rate of 22.05 kHz and a resolution of 16 bits/sample. The

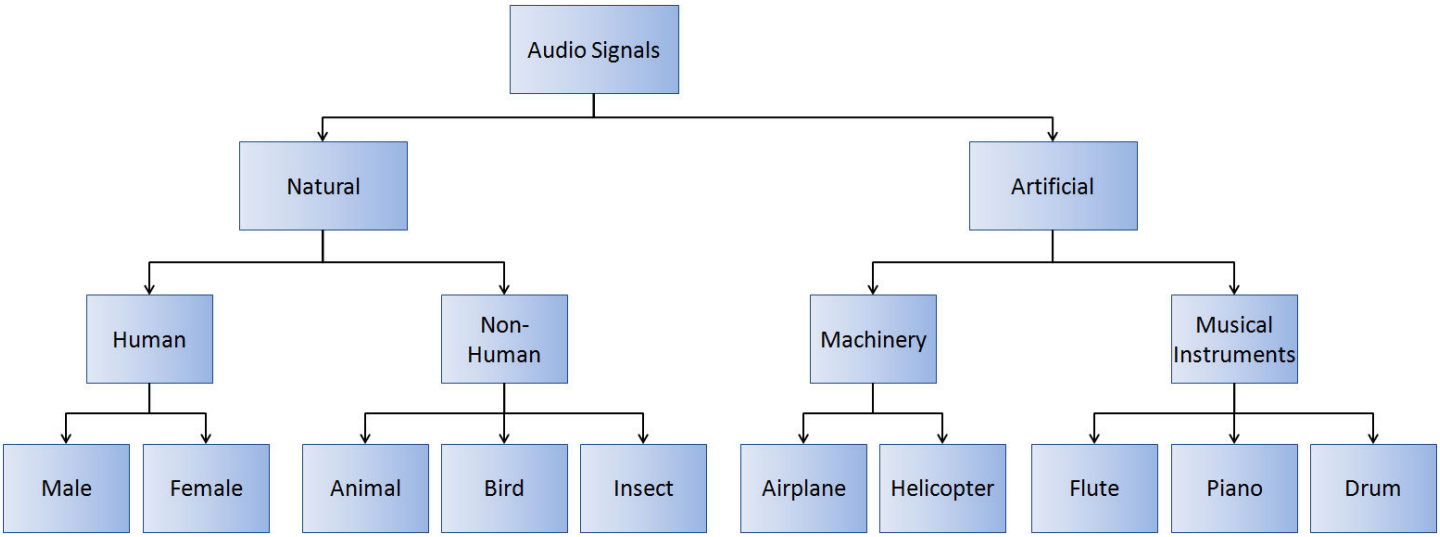


Figure 6.10: Organization of audio signals used in this work.

arrangement of this database is shown in Fig. 6.10. It is designed to have 10 different classes including 20 aircraft, 17 helicopters, 20 drums, 15 flutes, 20 pianos, 20 animals, 20 birds and 20 insects, and the speech of 20 males and 20 females. Most of the music samples were collected from the Internet and suitably processed to have uniform sampling frequency and duration.

The signal duration of 3s was utilized using the following rationale that the longer the audio signal was analyzed, the features better exhibited accurate audio characteristics. As the TFM feature extraction algorithm does not need any segmentation, theoretically there is no limit for the signal length. However, considering the hardware limitations of the processing facility, we are required to limit the duration of the signal.

6.6.4 Results and Discussions

We performed an audio classification as follows: 1) First, all the 192 audio signals are transformed into MP-TFD. 2) Next, NMF with decomposition order of 15 ($r = 15$) decomposes each TFM into 15 base and coefficient vectors. In the present study, experimentally, $r = 15$ is found to be a suitable choice for our application. 3) Then, 20 features (MFCC$_{1,\ldots,13}$, S$_h$, D$_h$, MO$_h$, S$_w$, D$_w$, MOw, MP) are extracted from each base and coefficient vector. 4) Finally, the extracted feature sets for the

signals are fed to the classifier based on LDA. Ten-group classification is performed (aircraft, helicopter, drum, flute, piano, male, female, animal, bird, and insect).

The performance of the proposed features for audio classification is evaluated through the following experiments:

Multiclass Classification

The extracted feature sets for the entire 192 signals were fed to the classifier based on LDA. Ten-group classification was performed. Table 6.5 shows the classification accuracy for different classification procedures. In this table, the first column shows the ten classes in the database and the number shows the number of signals in each class; for example, {'Aircraft'} includes 20 audio signals collected from different aircrafts. The second column in Table 6.5 shows the multiclass classification accuracy with regular LDA. By regular LDA, we mean that 75% of the signal samples in each class are used to train the LDA classifier, and the trained classifier is tested using the entire database. As can be seen in this table, helicopter, flute and piano donot have any misclassification. Drum, and male and female speeches achieve a classification accuracy of higher than 90%. The lowest accuracy rate belongs to animal, bird and insect sounds. The reason is that those classes are created from a large variety of creatures; for example, the animal class includes sounds of cow, elephant, hippo, hyena, wolf, sheep, horse, cat and donkey, which are very diverse in their nature.

Leave-one-out Classification

The accuracy rate of leave-one-out is shown in the third column of Table 6.5. The leave-one-out method is known to provide a least bias estimate. In the leave-one-out method, one sample is excluded from the dataset and the classifier is trained with the remaining samples. Then the excluded signal is used as the test data and the classification accuracy is determined. This is repeated for all samples of the dataset. Since each signal is excluded from the training set in turn, the independence between the test and the training set are maintained. The classification accuracy of 75.8% is acheived with the leave-one-out method.

Table 6.5: Classification Results - Features: The TFM Feature Extraction. Method: Regular - LDA, and Cross-Validated - LDA With Leave-One-Out Method

Class (#)	Accuracy(%)	
	Regular	Cross-Validated
Aircraft (20)	80	65
Helicopter (17)	100	88
Drum (20)	90	80
Flute (15)	100	90
Piano (20)	100	95
Male (20)	90	90
Female (20)	95	90
Animal (20)	55	45
Bird (20)	70	55
Insect (20)	75	60
Total (192)	85.5	75.8

One-against-all Classification

To further compare the proposed features, we performed a one-against-all classification method. In one-against-all classification, we learn ten individual binary classifiers, each one to distinguish the instances of a single class from the instances of all other classes. Thus one-against-all classification provides important information about the discrimination of the features of each class among the features from all classes. The gray bars in Fig. 6.11 demonstrate the one-against-all classification accuracy rate for the TF features. In this experiment, a binary SVM classifier [130] with Polynomial kernel and C = 100 was chosen experimentally to be a suitable choice in our application. The training stage is applied to 50% of samples in each class, and the test stage is applied to the entire database. Each bar indicates the accuracy of one class versus to the other 9 classes; for example, the first bar ('Ai') corresponds to the two class SVM of {'Aircraft'} and the remaining: {'Helicopter', 'Drum', 'Flute', 'Piano', 'Male', 'Female', 'Animal', 'Bird', 'Insect'}. As can be observed in Fig. 6.11, the one-against-all classification rate using the proposed TF features is higher than 94% for all the ten classes. The average classification accuracy of 97% with the one-against-all method proves the robustness of the proposed TF feature extraction technique.

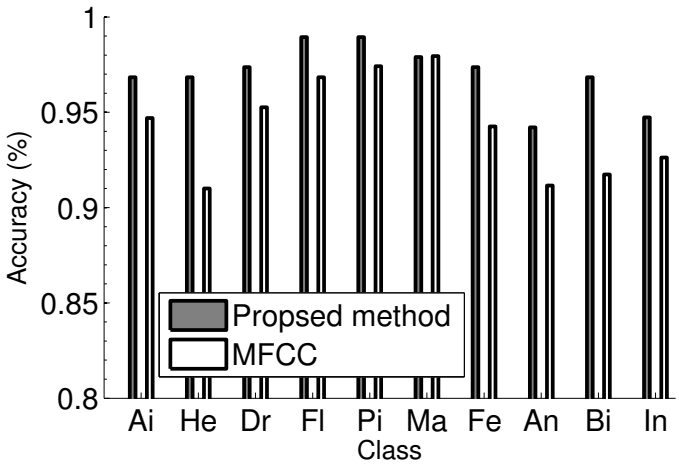


Figure 6.11: One-against-all Classification. The two letters in each class represents the first two letters of each class's name. Except for male speech class, the proposed long-term TF features improves the classification accuracies for all the audio classes with the highest increase (6%) in { 'Helicopter' } class, and the least increase (1.25%) in { 'Piano' } signals.

Comparison with MFCC Features

We compared the accuracy of the TFM decomposition features with the well-known MFCC features. MFCCs are short-term spectral features and are widely used in the area of audio and speech processing. In this paper, we compute the first 10 MFCCs for all the segments of the entire length of the audio signals and find the mean and variance of these 10 MFCCs as the MFCC features. For each audio signal we derive 20 features, 10 features are from the mean of the segment MFCCs and the remaining 10 are the variance of the segment MFCCs. These 20 features are computed for all the 192 signals and fed to a LDA-based classifier for classification. MFCC-based features results in a total accuracy rate of 74.5% which is 11% less than our proposed features. Additionally, we performed leave-one-out and one-against-all classifications to the MFCC features, and the classification accuracies of 67% and 94.3% were achieved, respectively. Comparing with the accuracies we obtained using TFM decomposition featues, 75.8% and 97% respectively, we can conclude that as we were expecting, the proposed long-term TF features are effective in characterizing the time-varying dynamics of environmental audio signals.

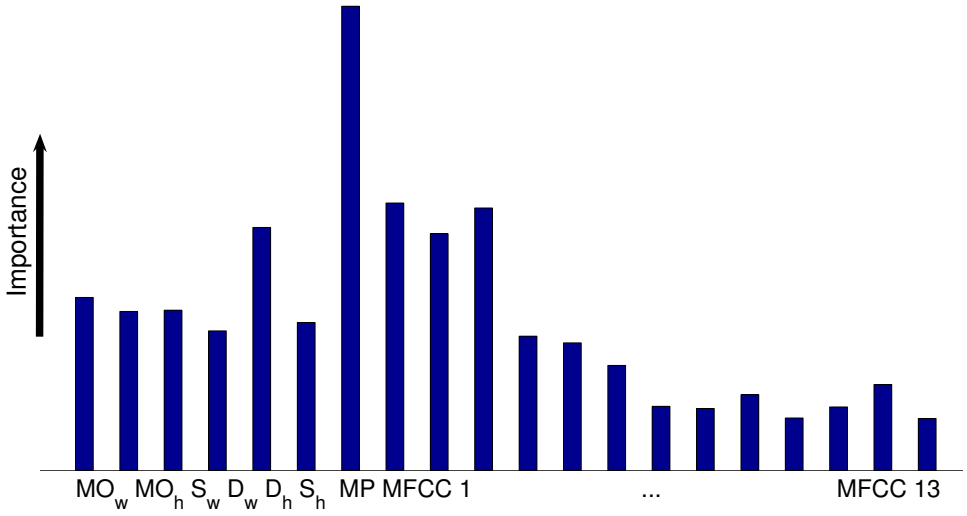


Figure 6.12: The relative height of each feature represents the relative importance of the feature compared to the other features.

Significant Features

Next, we evaluate the role of each feature in audio classification. To do so, we use the Students t-test to calculate the p value of the TF features. The feature with the smallest p value plays the most significant role in the audio classification. Fig. 6.12 demonstrates $\frac{1}{p\text{value}}$ as the relative importance of the 20 features we obtained in our proposed TFM feature extraction technique. As shown in this figure, the feature extracted from MP decomposition plays the most significant role in the classification accuracy. It can also be observed that the proposed TF features show a higher significance compared to the fourth MFCC feature and higher. This is proven by comparing the accuracy results with the TF features ($\mathrm{S}_h, \mathrm{D}_h, \mathrm{MO}_h, \mathrm{S}_w, \mathrm{D}_w, \mathrm{MO}w, \mathrm{MP}$) and with the MFCC coefficients only ($\mathrm{MFCC}_{1,\ldots,13}$). Using multiclass classification method, the TF features result in 68% accuracy which is significantly higher than the 55% accuracy that we achieved using only the MFCC features.

Spectrogram-based TF Features

In addition to the high classification accuracy, our proposed feature extraction methodology benefits from employing the MP-TFD as the TF representation plane. Because of the non-stationary dynamics of the real world signals, window-based TF approaches, such as Spectrogram, might loose the useful information of the signal. To demonstrate the accuracy of our claim, we perform our proposed feature extraction method using Spectrogram as the TF representation. To perform the proposed feature extraction methodology, first, Spectrogram of the 3 s duration of each signal is constructed. Next, NMF with decomposition order of 15 ($r = 15$) is performed on each Spectrogram. 19 features ($MFCC_{1,\ldots,13}$, $\mathrm{S}_h, \mathrm{D}_h, \mathrm{MO}_h, \mathrm{S}_w, \mathrm{D}_w, \mathrm{MO}w$) are extracted from each base and coefficient vector. Using the LDA multi classifier, we obtain classification accuracy of 79%, which is 4.5% more than the classification accuracy using MFCC features, and 6.5% less than the TF features derived from the MP-TFD. Through this experiment, it can be observed that i) the MP-TFD is more successful in characterizing the audio signals compared to Spectrogram; and ii) even when we are not using MP-TFD for the TF analysis, the TFM feature extraction methodology extractes long-term TF features that better represent the non-stationary structures of audio signals compared to the MFCC features which are obtained from short segments of a signal.

Noise Analysis

In the MP decomposition process, we extract the most coherent structure of a signal. Hence, it is expected that MP performs an automatic denoising, which allows the derived features from the MP-TFD to be robust to the noise in the audio signal. In order to evaluate the resilience of the proposed TF features, we train the audio classifier using clean audio database, but test the classifier using noisy audio signals. To make the noisy signals, we add a Gaussian random noise to the entire database, and instead of the features from the clean signal, we derive the features from the noisy signal. Our experiment showed that the MP-TFD based features are robust to noise with SNR of 10db or higher, while the features extracted using Spectrogram TF plane are unable to successfully classify the audio when SNR is less than 50db. This experiment proves that the MP-TFD is more robust to the presence of noise in signals, and as a result the audio classification is more practical

in the low SNR speech signals.

6.6.5 Summary

In this section, we applied our novel TFM feature extraction methodology to extract TF features for the purpose of environmental audio classification. Our proposed features were derived through three stages: First, signals were transformed into their TF representations using MP-TFD technique. Second, NMF technique was applied to the TFM of each signal, and decomposed the TFM into its significant spectral and temporal components. Third, a set of novel features ($\mathrm{S}_h, \mathrm{D}_h, \mathrm{MO}_h, \mathrm{S}_w, \mathrm{D}_w, \mathrm{MO}w, \mathrm{MP}$) were extracted from each decomposed vectors. These new features were combined with the traditional MFCC based features of each decomposed component in an attempt to improve the performance of the audio classifier.

The extracted features were evaluated using classification methods including: multiclass, leave-one-out and one-against-all classifications. The overall classification accuracies using these techniques was achieved 85%, 75.8%, and 97%, respectively. Furthermore, we compared the proposed features with the MFCC features. The accuracy rates achived using multiclass, leave-one-out and one-against-all classifications were 74.5%, 67%, and 94.3%, respectively. Additionally, the significance of the TFM features were demonstrated as the $\frac{1}{p\text{value}}$ obtained from the Students t-test. We showed that the feature extracted from MP decomposition plays the most significant role in the classification accuracy. We also observed that the proposed TF features showed a higher significance compared to the fourth MFCC feature and higher. Moreover, we investigated the advantage of MP-TFD to spectrogram in the proposed TFM decomposition feature extraction technique through two experiments as follows: in the first test, we used spectrogram to obtain the TFM of each signal, and an accuracy rate of 79% was obtained, which was 6% lower than the accuracy rate of the features obtained using MP-TFD. In the second experiment, we compared the classification of noisy signals using spectrogram with MP-TFD. The results showed that the MP-TFD based features were robust to the SNR of 10 db or higher, while spectrogram based features were robust to the SNR of 50 db or higher.

6.7 Experiment2: T wave Alternans Detection

In Chapter 3, we presented a novel Adaptive TF quantification that accurately quantified the known pattern in T wave alternans (TWA) signals to risk stratify patients with heart disease who may experience sudden death from ventricular arrhythmias [1]. Accurate estimation of TWA magnitude is important since larger TWA magnitude is associated with a higher risk of sudden cardiac death [131]. However, due to multiple periodic and non-periodic noise sources such as movement, and respiration in ambulatory ECG recordings, accurate measurement of TWA magnitude is technically challenging. In these cases where accurate quantification of TWA magnitude is not possible, we are interested in techniques that can accurately detect the presence of TWA. Therefore, in this chapter we use the developed pattern detection system as a complementary technique to detect TWA.

Once we construct the average Adaptive TFD of the aligned T waves, we consider the TFD as a matrix. This matrix combines the TWA matrix (**TWA**) and all the other non-desired components (**K**) as follows:

$$\mathbf{V} = \mathbf{TWA} + \mathbf{K} \tag{6.21}$$

where **V** is the average Adaptive TFD of the aligned T waves, **TWA** represents the TWA, and all the non-desired components are grouped into the noise matrix **K**. If we separate the T wave matrix from noise (**K**), and use the separated matrix to detect the TWA signal, we are able to detect the presence of TWA with a higher accuracy. We showed in this chapter that NMF decomposes a TF matrix into its spectral components with a high localization property, and applied NMF to TF matrix (TFM) decomposition of speech signals to automatically identify and measure the speech pathology problem. In this study, we use NMF to separate the TWA from the noise components, and refer to the proposed technique as NMF-Adaptive SM.

6.7.1 NMF-Adaptive SM

The schematic of NMF-Adaptive SM is displayed in Fig. 6.13.

[1] A comprehensive study of TWA background was presented in Section 3.5

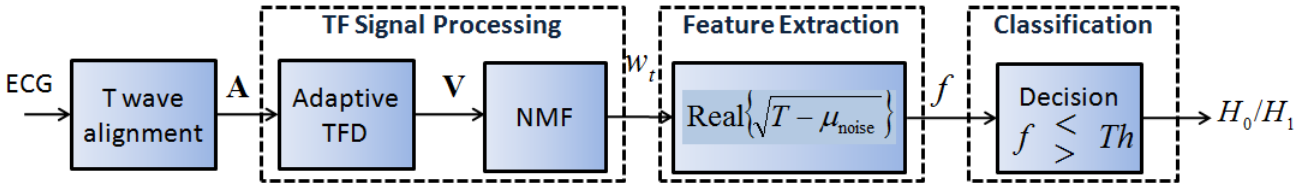


Figure 6.13: Schematic of the NMF-Adaptive SM for TWA detection.

Feature Extraction

The procedure is explained as follows:

- As shown in this figure, once we construct the average Adaptive TFD of the aligned T waves as shown in Eqn. 3.15, we consider the TFD as a matrix. This matrix combines the TWA matrix (**TWA**) and all the other non-desired components (**K**) as explained in Eqn. 6.21. In this approach, we focus our attention to separate the TWA from the noise in the spectral area that might dilute the presence of TWA. Therefore, rather than the whole matrix, we apply the NMF method to the last l rows of the matrix, **V**, where l corresponds to the number of samples in the spectral bandwidth of 0.36 to 0.5 cpb. We apply NMF to the matrix $\mathbf{V}_{l\times M}$, and decompose the TFM into two matrices, **W** and **H**:

$$\begin{aligned}\mathbf{V}_{l\times M} &= \mathbf{W}_{l\times r}\mathbf{H}_{r\times M} \\ &= \textstyle\sum_{i=1}^{r} w_i h_i\end{aligned} \tag{6.22}$$

  where r is the order of the decomposition, w_i is the ith column of the matrix **W** and h_i is the ith row of the matrix **H**. In this study, since the window length is chosen to be 64 ($M = 64$), l is equal to 10, and experimentally, $r = 3$ is found to be a suitable choice for our application. NMF estimates matrices **W** and **H** in a way that the columns of matrix **W** contain the spectral components presented in the TFM, and the rows of matrix **H** contain the corresponding temporal location of each spectral structure in the TFM.

- Next, we use Eqn. 3.18 in Chapter 3 to calculate the TWA magnitude of each spectral component. The component with the highest TWA magnitude is denoted as w_t, and Eqn.

6.22 is written as follows:

$$\begin{aligned} \mathbf{V} &= w_t h_t + \textstyle\sum_{\boldsymbol{I}} w_i h_i \\ \boldsymbol{I} &= \{i = 1, ..., r\} - \{i = t\} \end{aligned} \tag{6.23}$$

Due to NMF nature to represent all the components with the same spectral behavior in one column, we intuitively assume that the spectral magnitude of the TWA will be concentrated in one column (represented by w_t). Comparing Eqn. 6.23 with Eqn. 6.21, we conclude that using NMF on TFM of the aligned ST-T waveform, we are able to separate the **TWA** $= w_t h_t$ from the undesired ECG components that are caused by biological noise or a possible T wave alignment error. Therefore, in Eqn. 6.23, we use the w_t vector as a representative feature of the TWA present in the signal. The moment we apply the NMF method to decompose the TFM, the true value of the TWA is missed. We use the w_t vector as a better representation of the TWA.

- Finally, we derive the NMF-Adaptive SM feature vector as given below:

$$f_{NASM} = \{w_t, \alpha\} \tag{6.24}$$

$$\alpha = \text{Real}\left\{\sqrt{(T - \mu_{noise})}\right\} \tag{6.25}$$

where w_t is the decomposed spectral component with the highest TWA magnitude, T is the magnitude of that component at 0.5 cpb, μ_{noise} is the noise estimate calculated from w_t at the spectral bandwidth, 0.44 to 0.49 cpb, and l is the length of w_t.

Classification

Once features are extracted as in Eqn. 6.25, the features are fed into a linear discriminant analysis (LDA) in order to train a classifier to detect the TWA signals. Then, we use the trained classifier to detect TWA in subsequent new ECG signals.

6.7.2 Example: Synthetic Signal

The application of our algorithm to quantify TWA in ambulatory ECG recordings is shown in Fig. 6.14. Here we simulate TWA by adding a rectangular pulse of 5μV to alternate beats. The aligned

T waves are shown in Fig. 6.14(a). Fig. 6.14(b) illustrates the reconstructed TFM, $\mathbf{V}_{l\times M}$, where $l = 10$ corresponds to the point 0.36 cpb. The decomposed matrices, **W** and **H**, resulting from NMF with decomposition order three ($r = 3$) are shown in Figs. 6.14(c) and (d), respectively. As evident in these figures, the columns of matrix **W** represent the components present in the TFM of the T waves, and the rows of matrix **H** show the location of each corresponding component. We calculate TWA magnitude for each decomposed component using Eqn. 3.18, and choose the component with the highest TWA magnitude. As shown in Fig. 6.14(c), the third component, which is indicated by a dashed box, has the greatest T wave variation. Fig. 6.14(e) illustrates the TFM represented by the third component (**TWA** $= w(3)h(3)$), and Fig. 6.14(f) shows the TFM represented by the rest of the components (**K** $= w(1)h(1) + w(2)h(2)$). From the decomposed matrix (**TWA**) in Fig. 6.14(e), we observe that NMF-Adaptive SM successfully distinguishes the TWA energy at 0.5 cpb which is masked by the noise in the original TFM ($\mathbf{V}_{l\times M}$).

6.7.3 Experiment: Real Ambulatory ECG Signals

Dataset

Real world ECG recording with inherent noise were obtained from 26 normal subjects who underwent 2 channel ambulatory ECG recording (GE Healthcare, Inc.) for 24-48 hours duration at our institution. The ECGs were recorded at a sampling rate of 125 Hz and then exported for custom analysis. Each ECG channel was included as a separate record. Baseline correction and QRS onset annotations were performed as described previously. The noise level of the recordings was determined as the standard deviation (SD) over the first 80 ms of the TP interval after correcting baseline wander. As with the synthetic ECG recordings, a simulated TWA signal of 5 μV was added to the ECG. This was achieved by increasing T wave amplitude of even beats and decreasing T wave amplitude of odd beats uniformly across the T wave from a point 40 ms after QRS offset to the end of the T wave. Hence two groups of ambulatory ECGs were generated, one without simulated TWA (ie TWA magnitude = 0 μV) and the other with simulated TWA (ie TWA magnitude = 5 μV).

Results

In order to evaluate TWA detection, NMF-Adaptive SM was then performed on the first 64-beat segment of each ambulatory ECG channel. We pre-specified a TWA detection threshold of 5 μV as this cutpoint approximates the TWA magnitude measured by Klingenheben et al [131] in patients with heart disease using a similar definition of TWA as our study. NMF-Adaptive feature set (Eqn. 6.25), the measured TWA and the K_{score} of the SM, and the measured TWA using MMA were calculated from 64-beat segment of each ambulatory ECG group. The extracted features were fed into an LDA classifier in order to group the ECG segments either with and without TWA. In order to compare the accuracy of the methods for detecting TWA, receiver operating curves (ROC) were computed with the area under the curve indicating relative TWA signal discrimination (Fig. 6.15). The area under the ROCs for NMF-Adaptive SM and SM were 0.92 and 0.77, respectively, indicating better TWA discrimination with NMF-Adaptive SM compared to SM.

In clinical medicine, the absence of the TWA signal at certain heart rate intervals is relevant because the risk of adverse cardiac events in patients with heart disease is sufficiently low that the treatment is not necessary [132]. Therefore, it is important to preserve 100% specificity with a TWA detection algorithm such as NMF-Adaptive SM. Based on the ROCs, the maximum sensitivity for TWA detection while preserving 100% specificity is 48% for NMF-Adaptive SM and 20% for the the K_{score} of the SM, which represents an 140% improvment over the conventional TWA detection algorithm. From the ROCs, we may conclude that NMF-Adaptive SM performs significantly better than SM and MMA, and provides a better statistics for TWA detection. Table 6.6 provides the classification accuracy for NMF-Adaptive SM, SM, and MMA. As predicted from the ROC plots, NMF-Adaptive SM offered significantly higher true negative and true positive rates (87% and 91%, respectively) compared to the other two approaches. MMA resulted in a high true negative (92%), while the true positive was very low (58%). The results for SM were poor for both the true positive and negative rates.

Table 6.6: TWA Detection Rate

Method	Class	0μV TWA	5μV TWA
NMF-Adaptive SM	0μV TWA	87%	13%
	5μV TWA	9%	91%
SM	0μV TWA	73%	27%
	5μV TWA	37%	63%
MMA	0μV TWA	92%	8%
	5μV TWA	42%	58%

6.7.4 Summary

In this section, we applied the proposed TFM feature extraction methodology to extract TWA features from ambulatory ECG signals, and called the new method NMF-Adaptive SM. This method considered the TFD of the aligned T wave signals as a matrix, and applied NMF to separate the TWA components from the other TF components in the TFD. It then used the separated TWA component to detect the presence of TWA at 0.5 cpb. We applied NMF-Adaptive SM to detect TWA in ambulatory ECGs. The area under the ROC for NMF-Adaptive SM, K_{score} of the SM and MMA was 0.92, 0.77 (p ¡ 0.001 NMF-Adaptive SM vs. K_{score} of the SM) and 0.7, respectively, indicating a superior detection accuracy of the proposed NMF-Adaptive SM to SM and MMA.

6.8 Chapter Summary

The highlighted blocks in Fig. 6.16 display the contribution of this chapter. The chapter calibrated the proposed TFM quantification methodology to further improve the representation of the extracted TF features. Non-negative matrix factorization (NMF) was selected as the desirable matrix decomposition (MD) technique for TFM feature extraction as proposed in Chapter 5. The NMF-based TFM quantification approach accurately characterized a given TF plane with highly representative and localized TF features. Other contributions of the present work included integration of MP-TFD algorithm with NMF optimization to seed the decomposed matrices in NMF optimization. This integration improved the time convergence of the algorithm to be half the time required in using the randomly seeding technique. The proposed technique did not have to take

any stationary assumption about the signal. Instead, NMF adaptively decomposed the TFM of long-term signals into segments with similar spectral structures. The features extracted from these segments successfully quantified signals' TF structures, and provided an efficient and reduced TF representation of the signal.

In this chapter, we applied the developed TFM feature extraction to two real-world applications: Audio and TWA classifications, in which several experiments were performed to evaluate the proposed framework. The high accuracy resulted under noisy or non-stationary conditions verified the effectiveness of the proposed TFM quantification methodology compared to the conventional techniques. A significant outcome clearly demonstrated the potential of the new technique as a true non-stationary tool to identify unknown and complex patterns in real-world data, such as environmental audio classification and TWA detection. The next chapter covers the process of developing a novel discriminant feature clustering from the obtained TFM features, and its utility in performing discriminative signal classification.

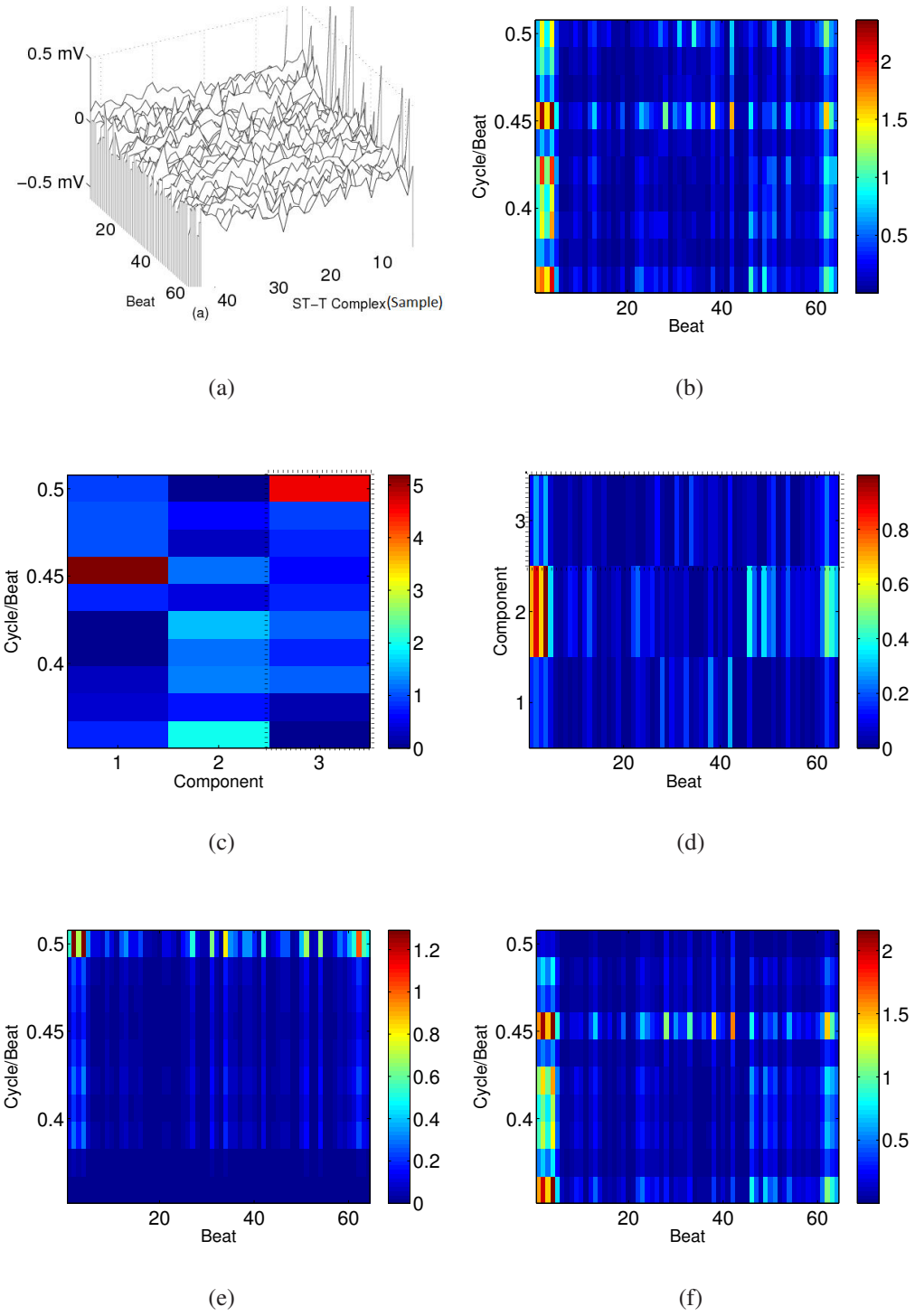


Figure 6.14: NMF-Adaptive method is demonstrated. (a) The aligned T waves for a 64 beat ECG segment. (b) The average Adaptive TFD of the aligned T waves. (c) The decomposed spectral components. (d) The decomposed temporal components. (e) The TWA matrix separated from the TFD. (f) The undesired part of the TFD (**K**). As can be seen, NMF-Adaptive separated the TWA energy at 0.5 cpb from the original TFM.

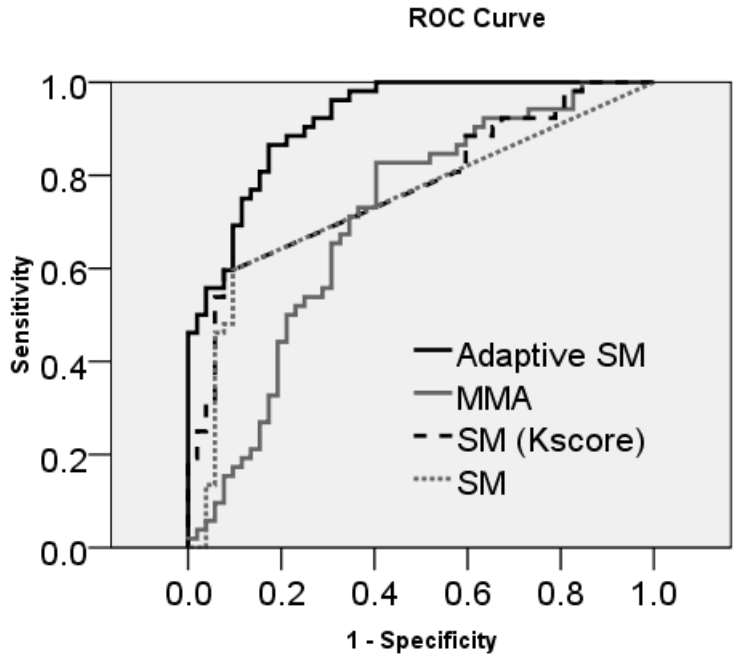


Figure 6.15: Receiver operating curves for the SM and NMF-Adaptive SM methods are plotted. In this analysis, ambulatory ECGs without added TWA are considered negative, while the ECGs with added TWA of 5μV are considered positive. The area under the ROC for NMF-Adaptive SM and K_{score} of the SM are 0.92 and 0.77, $p < 0.001$, respectively. The area under the ROC for SM and MMA are 0.74 and 0.7, respectively.

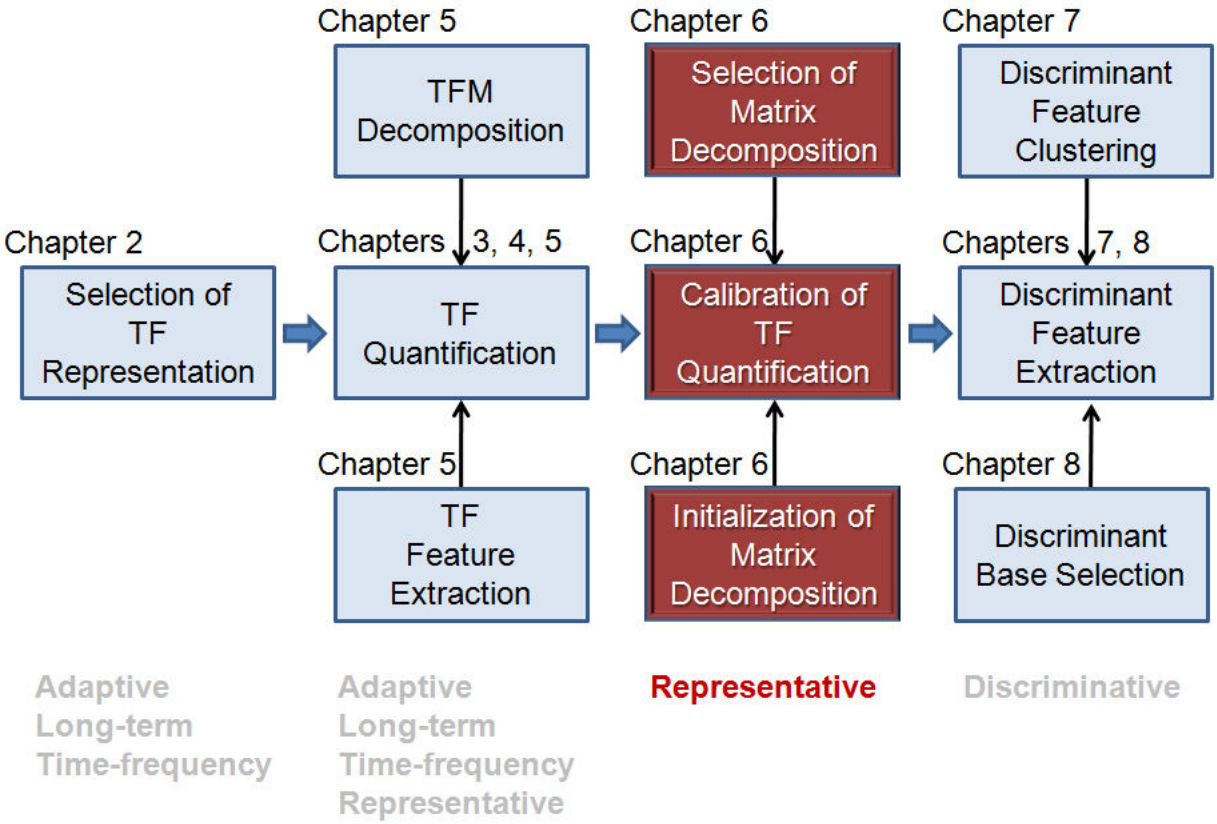


Figure 6.16: Flowchart of the proposed contributions.

Chapter 7

DISCRIMINANT FEATURE CLUSTERING

7.1 Motivation

OUR goal to achieve a higher accuracy pattern classification system in Chapters 5 and 6 motivated us to develop the TFM quantification technique. This method extracts unique TF features in a way that they effectively represent any given signal in the feature space. An example of such an ideal feature space is shown in Fig. 7.2(a). This figure shows a desirable situation in which the signals from two classes (i.e. class A and class B) are separable in the feature space. The features are so well-defined that we can easily distinguish any signal in class A from the signals' in class B, and vice versa. However, in most real-world scenarios, the features from separate classes tend to have an overlap in the feature domain. The reason for such overlapping includes: (i) The non-stationarities in the real-world signals cause variations so that the obtained features may show a spread over the feature space, which cannot fit in the same category. (ii) In most of the real world applications, the nature of signals from different classes are very similar. There are only slight variations in the signals' patterns from one class to another class. Since the extracted features represent the general characteristics of each given signal, they may not necessarily represent the discriminating structures in each class. As a result, an overlap occurs in the feature space.

Fig. 7.2(b) displays a two-class feature space with overlapping features. In this example, the features of classes, A and B, are shown with two different markers (i.e. cross and asterisk). As

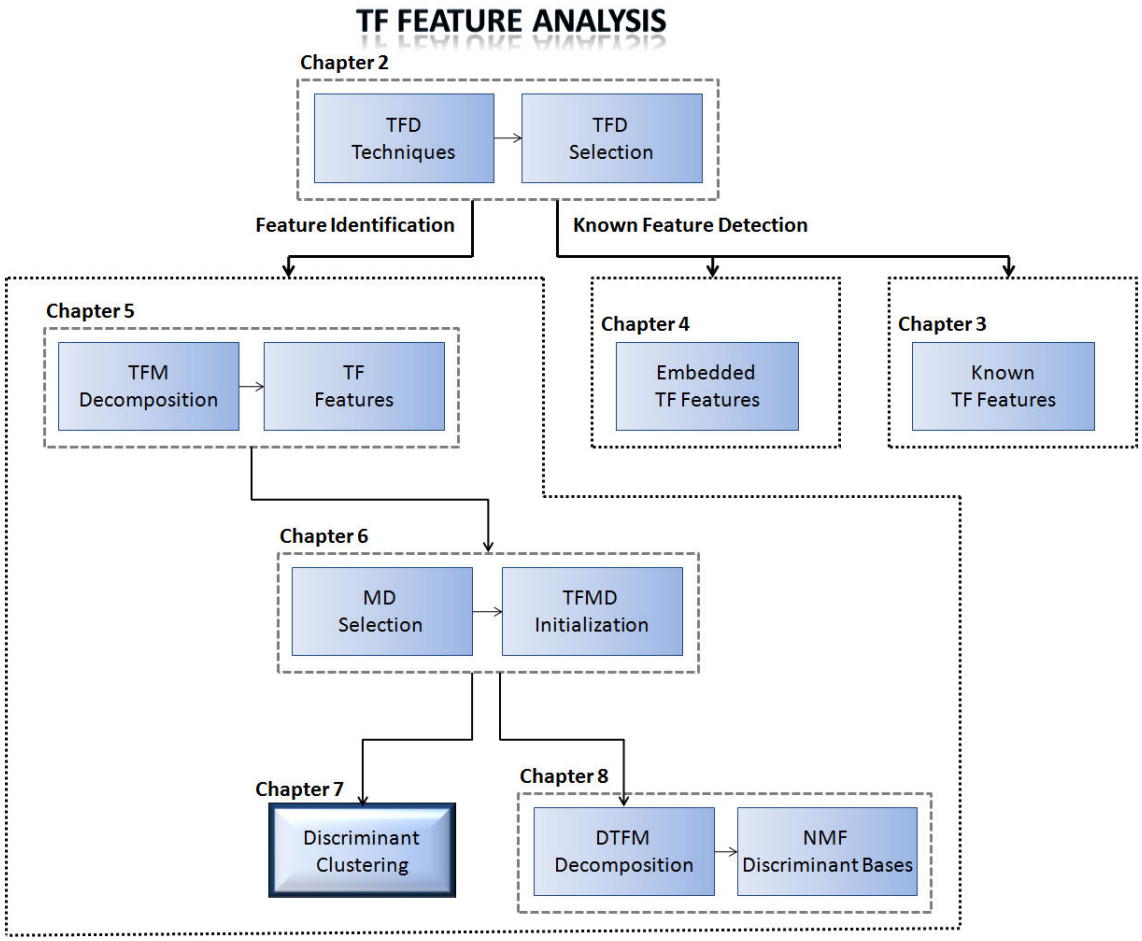
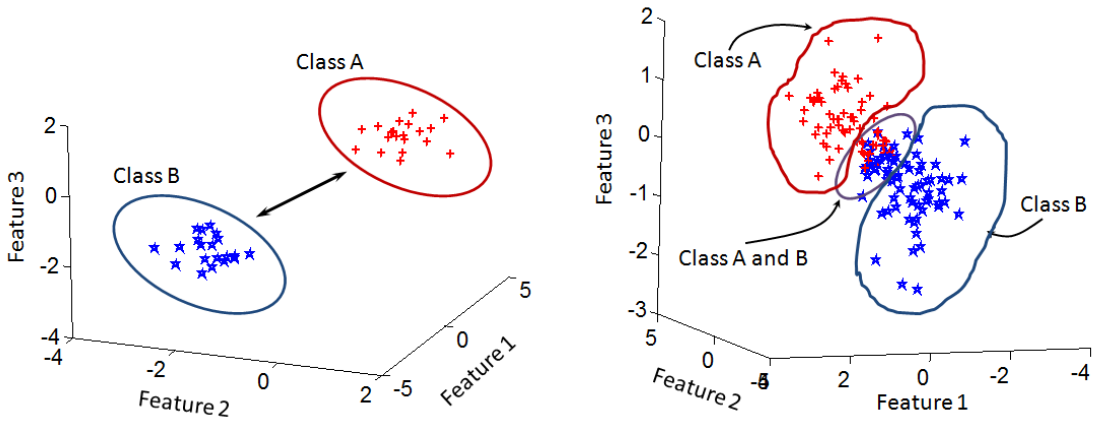


Figure 7.1: Chapter 7 - Clustering of discriminant features.

seen in this figure, the majority of class A and B features are separate from each other. However, there are some samples that overlap in this feature domain, which belong to the common structure between the two classes. The overlapping features cannot distinguish the discriminating structures of the signals, so they degrade the classification performance.

7.1.1 Proposed Contribution

In order to discriminate different classes in the feature domain despite the features' overlapping, complex learning algorithms such as artificial neural networks (ANN) [127] have been developed in the literature. While these techniques tend to increase the classification rate, they are abstract



(a) (b)

Figure 7.2: (a) An ideal scenario in which features are separate in the feature domain. (b). A scenario where features overlap in the feature space.

approaches, which do not provide any insight about the signals' structures, and are also computationally expensive.

As our goal to select the extracted features to represent the discriminative structures between signals from different classes, we focus on developing a new framework which we call "discriminant feature clustering". In this system, we first identify the features that represent the non-overlapping areas as discriminant signatures, and then train a simple classifier based on the selected features. Since these features better represent the discriminative structures in each class, the trained classifier will achieve a higher accuracy rate compared to a classification system trained without any discriminant feature clustering. In order to proceed towards such a feature clustering, we need a clustering technique that labels the features according to their location in the feature space. Such a clustering technique divides the features similar to what we manually performed in the example shown in Fig. 7.2(b). The class A feature points, which are far from the feature points of class B, are identified as the discriminating pattern in class A, and vice versa.

The discriminant feature clustering methodology in this work is a fusion of an unsupervised classifier and a supervised labeling, in which the unsupervised learning clusters the features, and the supervised labeling identifies the discriminant clusters. Fig. 7.3 displays our contributions

in this chapter. As seen, we select k-means and self organizing tree map (SOTM) as the cluster analyses methods. Next, we employ a clustering analysis technique to cluster the features, and then, we propose a new cluster labeling approach to identify the discriminative clusters. We call the proposed labeling technique ”supervised cluster labeling” as the technique is performed based on the known labels. Two labeling methods are proposed: hard and fuzzy labellings. Finally, we apply the developed system to two real-world applications.

Unsupervised clustering has been widely studied in literature; however, to our knowledge this chapter is one of the first known works that studies the clustering approaches for identification of discriminant features. In this work, the already available clustering approaches allow for extraction of underlying characteristics of the data, and then the proposed supervised labeling is used to interpret the discovered clusters.

7.2 Methodology

Fig. 7.4 displays the schematic of our proposed discriminant feature clustering method. As can be seen in this block diagram, once the features are extracted, we perform an unsupervised clustering technique to the entire feature set. Next, we label the features based their settings in the feature domain. The following examples demonstrate the proposed methodology and its advantageous to summarize key features of the data.

Fig. 7.5 shows a synthetic example. In this example, one normal and one abnormal signal is generated using Eqn. 8.16. The normal signal consists of seven frequency modulated components. As descriptor of the signal under abnormality conditions, three of the components are transformed into transients. Figs. 7.5 (a) to (d) display the generated normal and abnormal signals in time and TF domains, respectively. Employing the proposed TFM quantification, the TF features of the signals are extracted, and are depicted in Fig. 7.5(e). As it can be seen in this figure, because the normal and abnormal signals contain a similar structure, not all of the features points are separated in the feature domain. Only three of the feature points which belong to the transient components in the abnormal signal are far from the rest of the feature points.

Our proposed feature clustering system is performed in two stages: In the first stage, an un-

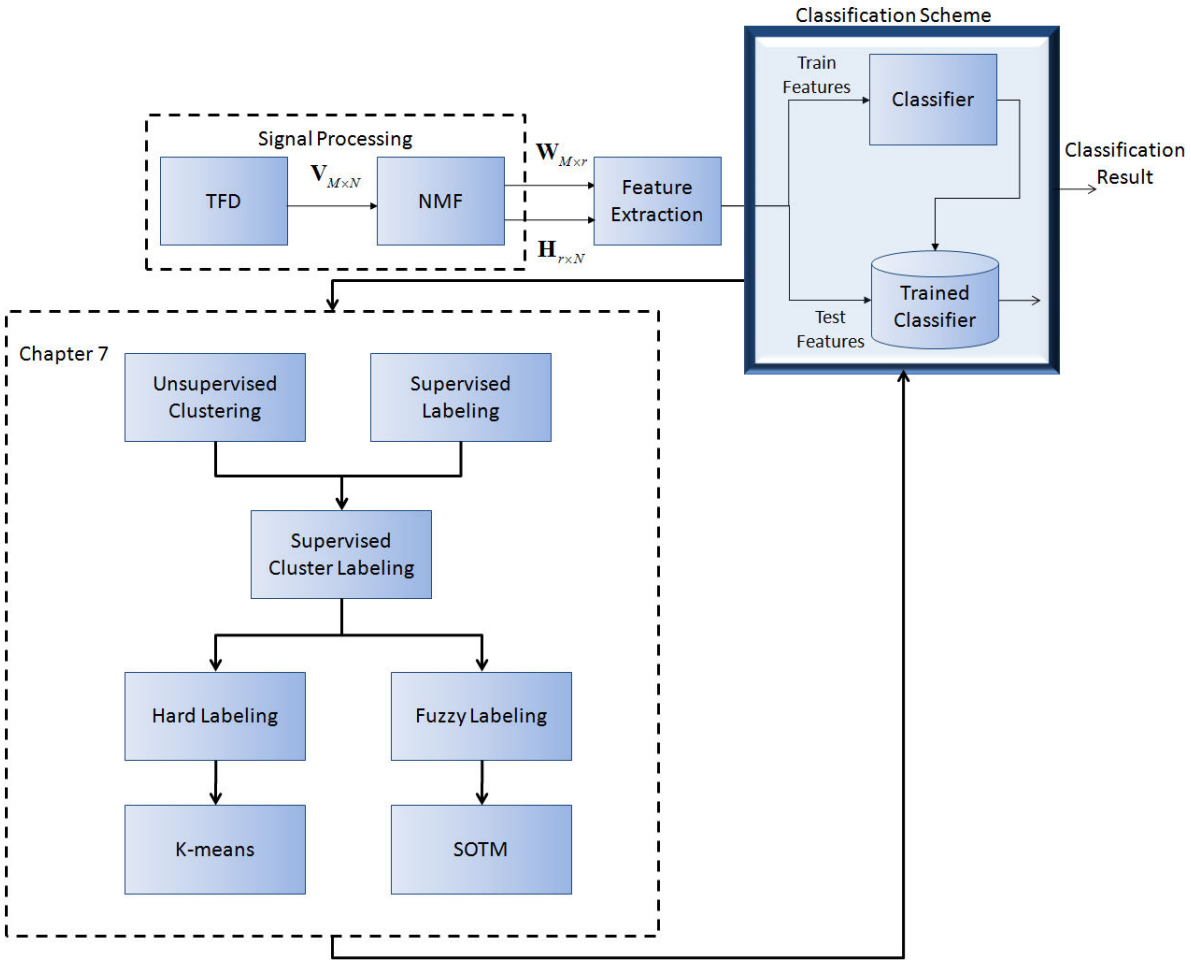


Figure 7.3: Chapter 7 - Discriminant Feature Selection.

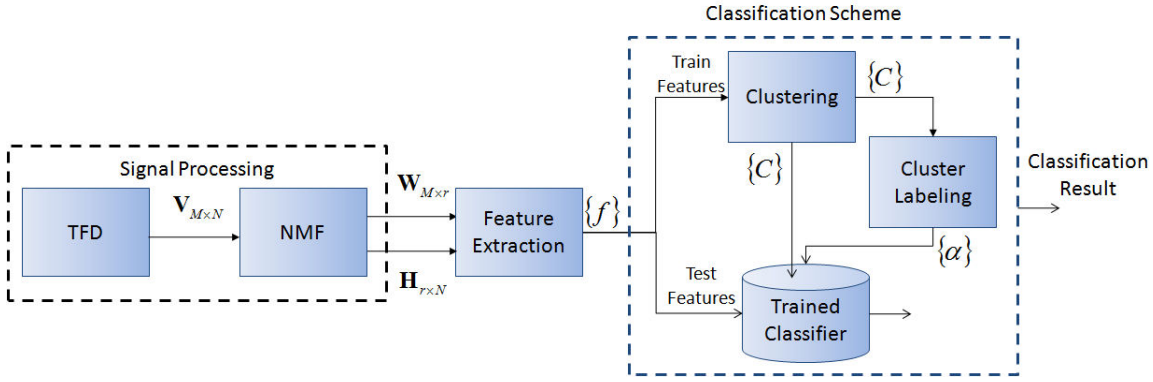


Figure 7.4: The block diagram of our proposed method for discriminant feature selection.

supervised learning is performed on the entire features (i.e., both normal and abnormal) to identify the clusters in the feature space. In the second stage, a supervised labeling decides whether each cluster represents abnormality structures. Fig. 7.5(f) displays the two identified clusters in this example. The cluster, which consists of abnormality features, is labeled as the discriminant structure corresponding to the abnormality pattern, and the other cluster is labeled as the features corresponding to the frequency modulated components which exist in both normal and abnormal signals.

As demonstrated in the above example, the fusion of unsupervised clustering and supervised labeling enabled us to identify the discriminant features that belong to the transients in the abnormal signal. The next section explains unsupervised clustering techniques.

7.3 Clustering Techniques

Supervised and unsupervised learnings are two classification approaches where in the former method, the data labeling is known, and in the latter one, there is no information available about the data labels. As explained in Chapter 1, unsupervised learning enables us to obtain more adaptive and meaningful classes, which correspond to the natural characteristics of the data. Unsupervised cluster analysis is the foundation of any unsupervised classifier which has been used to organize a large dataset. Clustering techniques have been utilized in areas such as data mining [133], information retrieval [134, 135] image segmentation [136], signal compression and coding [137] and machine learning. While the widespread application of unsupervised clustering includes the cases where labeling a large and complex dataset can be surprisingly costly, their application in the area of signal feature selection is relatively unexplored [138].

Numerous clustering algorithms have been proposed in the literature and new ones continue to appear. Classification of the existing clustering methods can be performed based on different measurement criteria. Typically the algorithms can be divided into two major classes: parametric and non-parametric.

- Parametric cluster estimation: This approach is based on the assumption that the underlying data density is a mixture of K Gaussian clusters. Gaussian Mixture Model (GMM) [139] is

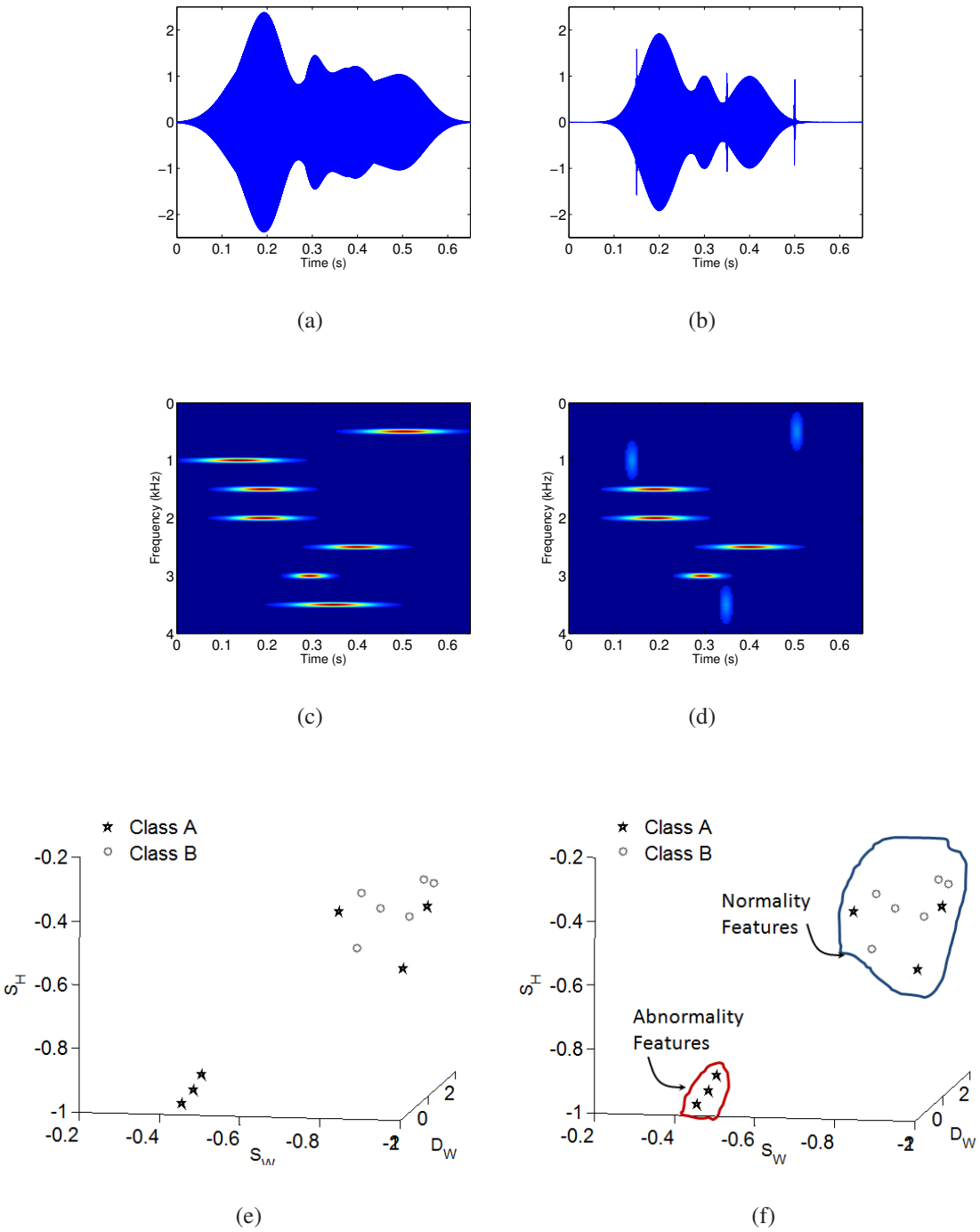


Figure 7.5: (a) Normal synthetic signal. (b) Abnormal synthetic signal. (c) TF representation of the normal signal. (d) TF distribution of the abnormal signal. (e) Feature space. (f) Supervised cluster labeling identified the abnormality features.

a well known parametric density estimation method that is used for partitioning a given data. The major problem with this approach is that if the data do not conform to the assumptions made by the technique then the imposed structure on the data may not disclose the true structure.

- Non-parametric cluster estimation: Non-parametric approaches utilize methods of scale-space filtering to extract robust structures within a data set. There are two non-parametric approaches: Partitional clustering and Hierarchical clustering.
  - Hierarchial: Hierarchical algorithms are mostly based on the agglomerative hierarchical clustering. These algorithms attempt to organize data in a nested sequence of groups which can be displayed in the form of a dendrogram or a tree. The major drawback of hierarchial approach is that the entire dendogram is sensitive to previous cluster merging, i.e.data are not permitted to change cluster membership once assignment has taken place.
  - Partitional: These types of methods typically start with a data partitioning into a small number of clusters and increase the number of partitions into which the data is divided. The precise partitioning is performed so as to minimize or maximize some objective function. Data are free to change so the partitioning membership. A precise choice of the objective function plays an important role on efficient partitioning of the data.

Partitional clustering algorithms attempt to obtain a single partition of the data. These methods have the advantage in applications where a large amount of data is to be processed. In such cases, the use of a dendrogram is not computationally feasible. The partitional techniques usually generate clusters by optimizing a criterion function which is defined either locally or globally. The algorithm is run multiple times with different starting points and the best configuration is then selected as the result of clustering. One of the most popular partitional clustering algorithms is k-means clustering algorithm which is explained in the following section.

7.3.1 k-means Clustering

The k-means clustering is one the simplest and the most popular unsupervised clustering algorithms. The algorithm is computationally efficient and is advantageous on a dataset that consists of compact and well separated clusters [140]. Given a feature set, $\{f\}_{1,\dots,Z}$, the following phases are performed in the algorithm [138]:

1. The method starts with K initial random centroids, $\{C_u\}_{u=1,\dots,K}$.

2. It classifies the feature samples into the nearest centroid according to the squared Euclidean distance (ED). To do so, it first calculates the squared ED of any given sample to all the centroids as given in the following equation:

$$\left\{e_z^2\right\} = \sum_{u=1}^{K} \|f(z) - C_u\|^2 \tag{7.1}$$

   Then, the algorithm assigns the sample to the centroid with minimum ED.

3. The mean of the points in each cluster is computed as the new cluster centroids:

$$C_u = \frac{1}{Z_u} \sum_{u=1}^{Z_u} f(z)^u \tag{7.2}$$

   where, Z_u is the number of feature samples assigned to cluster u, and $\{f(z)^u\}_{z=1,\dots,Z_u}$ are the assigned samples to cluster u.

4. The algorithm iteratively repeats stages 2 and 3 unless the new cluster centers are the same as or close enough to the centroids of the previous stage.

Although k-means clustering has been successfully used in the literature for various applications, one of its drawbacks is that the number of clusters has to be known at the beginning of the process. In the next section, we explain a clustering technique that does not require any information about the number of clusters in the feature domain.

7.3.2 Self-organizing Map (SOM)

Self-organizing map (SOM) or self-organizing feature map (SOFM) is a type of Artificial Neural Networks (ANN) that is trained using unsupervised learning to project high-dimensional data into a low-dimensional discrete space, called a map. SOMs are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. SOMs are well-known data mining tools which are used for visualization and exploratory purpose. However, one of the main disadvantages of SOM is the nodes being trapped in the low density areas [141]. To overcome this drawback, self organizing tree map (SOTM) is proposed. Unlike the SOFM, SOTM does not suffer from the disadvantage of nodes being trapped in the low density areas [141] and the network has a dynamic structure that grows from a single node.

Self Organizing Tree Map (SOTM)

SOTM was first introduced in [142]. The algorithm maps the data from a high dimensional Euclidean feature space onto a finite set of prototypes. Like most of the clustering algorithms, it attempts to organize unlabeled feature vectors into the clusters in a way that all the samples within a cluster are more similar to each other than those of other clusters. Each cluster is then represented using one or more prototype. Unlike clustering methods like K-means where the number of clusters should be known beforehand, in SOTM the number of clusters is determined by the algorithm based on parameters, which define the desired resolution of the clustering. The steps involved in the SOTM algorithm are briefly explained below:

1. The weight vectors are initialized randomly $\{\gamma_u(t)\}_{u=1,...,K}$, where K is the number of clusters. The random value is usually a vector from the training set.

2. For a new input vector, the distance from the input vector and all of the existing nodes, d_u, is calculated as

$$d_u = \left\{ \sum_{z=1}^{Z} \left[f(z) - \gamma_u(t) \right]^2 \right\}^{1/2} \quad u = 1, ..., K \tag{7.3}$$

where $\gamma_u(t)$ is the node of the cluster u at time t.

3. Select the node with the minimum distance d_u as the winning node, u^*

$$d_{u*}(\vec{f}, \vec{\gamma}_u(t)) = \min d(\vec{x}, \vec{\gamma}_u(t)) \tag{7.4}$$

4. The minimum distance, $d_{u*}(\vec{f}, \vec{\gamma}_u(t))$ is then compared with $H(t)$, the hierarchical control function, which decreases over time. If the input vector is within the threshold $H(t)$ of the winning node, the weight vector is updated based on the following update rule:

$$\vec{\gamma}_u(t+1) = \vec{\gamma}_u(t) + \lambda(t)[\vec{f} - \vec{\gamma}_u(t)] \tag{7.5}$$

   Where $\lambda(t)$ is the learning rate, which decreases with time. When the input vector is farther from the winning node than the threshold, a new subnode is generated from the winning node at $\vec{f}$.

5. Checking the terminating conditions; The algorithm will stop of any of the following conditions are fulfilled

   - Maximum number of iterations is reached.
   - Maximum number of clusters is reached.
   - No significant change occurs in the structure of the tree.

   Otherwise the algorithm is repeated from step 2.

The Hierarchical Control Function acts as an ellipsoid of significant similarity. $H(t)$ can be assumed as a global vigilance threshold that is used for measuring the proximity of a new input sample to the nearest existing node in the network. Samples that fall outside the scope of the nearest existing node, result in generation of a new node as child of the winning node. By initializing $H(t)$ to start from a large value, the clusters discovered at the early stages of the clustering will be far from each other. Decay of $H(t)$ over time results in partitioning the feature space in low resolution at the early stages of the clustering, while favoring partitioning at higher resolutions later. There

are two standard hierarchical control function proposed for the original SOTM algorithm: linear and exponential decay.

$$\begin{aligned} H(t) &= H(0) - \left[\left(1 - e^{-\zeta/\tau H} H(0)\right)/\zeta\right] t, \\ H(t) &= H(0) e^{-t/\tau H}, \end{aligned} \tag{7.6}$$

where τH is a time constant, which is bound to the projected size of the input feature F, $H(0)$ is the initial value, t is the number of iterations (or sample presentation) and ζ is the number of iterations over which the linear version of $H(t)$ would decay to the same level as the exponential version. One benefit of initializing $H(t)$ to a large value, possibly larger than the maximum variation within the data, is that all levels of resolution across the data can be explored.

The learning rate in Eqn. 7.5, $\lambda(t)$ is an important factor in organizing the network. $\lambda(t)$ can operate in number of different global or local modes. In global modes a single learning rate is applied to all node, whereas in local modes an individual rate operates for each node a set of nodes. There are a few modalities proposed for the operation of the learning rate. Some of these modes are discussed [141].

7.4 Labeling Techniques

Assigning the right label to each cluster is one of the critical concerns in cluster labeling. We propose two methods to label the obtained clusters as explained in the following sections.

7.4.1 Method 1: Hard Labeling

In an E-class classification problem, say class 1, class 2, ..., class E, this method decides whether each cluster represents class 1, 2, ..., or E.

- First, the clusters are identified, say K clusters $\{C_1, C_2, ..., C_K\}$.
- Next, the number of each class feature vectors are calculated in each respective cluster as $\text{NUM}_1(u)$, $\text{NUM}_2(u)$, ..., $\text{NUM}_E(u)$ representing the number of classes 1 to E features in the uth cluster, respectively.

- Then, the class with the majority numbers defines the label of each cluster. The calculation is proceeded as shown in the following equation:

$$\begin{aligned} &\alpha_u = \text{Max NUM}_{\mathbf{e}}(u), \\ &u = 1, ..., K \end{aligned} \tag{7.7}$$

where α_u is the label defined for the uth cluster.

- Once the training stage is completed, the estimated clusters and the calculated labels, denoted with $\{\alpha_1, \alpha 2, ..., \alpha_K\}$ are passed to the test stage. In the testing stage, any new feature is classified based on which cluster it belongs to. First, the centroid of each cluster is calculated as the mean of the feature points in that cluster as shown in the following equation:

$$\text{Center}_u = \sum_{z=1}^{k_u} c_u(z) \tag{7.8}$$

where c_u is a feature in the uth cluster, and k_u is the number of features in that cluster.

- For each feature, we find the nearest cluster using Euclidean distance criterion. If the number of the feature vectors that belong to class e clusters is dominant, the signal is classified as a class e signal. To perform this calculation, for any new feature vector f, this procedure is performed as shown in below equations:

$$f \in \alpha_u \quad \text{if} \quad \alpha_u = \underset{u=1,...,K}{\text{Label min}} \left| f - \text{Center}_u \right|, \tag{7.9}$$

which defines the label of the uth cluster as the class of the feature f.

We call this procedure as hard labeling as each cluster is distinguished with one label.

7.4.2 Method 2: Fuzzy Labeling

- Our second proposed cluster labeling calculates the label of each feature as a membership matrix $\mathbf{M}_{K \times E}$, where each entry in this membership matrix, m_{ue} (which we call a membership coefficient) indicates the probability of a vector in the cluster u to belong to class e.

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1E} \\ m_{21} & m_{22} & \cdots & m_{2E} \\ \vdots & \vdots & \vdots & \vdots \\ m_{K1} & m_{K2} & \cdots & m_{KE} \end{bmatrix} \tag{7.10}$$

where the membership coefficients are calculated based on the distribution of each class in different clusters as shown in the below equation:

$$\begin{aligned} m_{ue} &= p(\theta_{\text{Class}_e}|\text{Cluster}_u) \\ &= \frac{\text{NUM}_{e(u)}}{k_u} \end{aligned} \tag{7.11}$$

where $\text{NUM}_e(u)$ is the number of features belong to class e that exist in cluster u, and k_u is the total of features in the uth cluster. These coefficients will be used in calculation of the membership degree for each of the test vectors. The main advantage of calculating the membership coefficients is to take into consideration the overlap of the classes in the feature space. When a signal is projected onto the feature space, some of its representing vectors may fall in the areas which are common within two or more classes. By using this approach, less weight is associated with the vectors that are located in the overlap area.

- In the test stage, each of the feature vectors representing a test signal is assigned to one the cluster centers found in the previous stage based on the minimum Euclidean distance criterion. For each signal the scatter vector, S is defined as

$$S = [s_1, s_2, ..., s_K] \tag{7.12}$$

where s_u is the number of the representing vectors for a test signal that fall within the uth cluster and K is the number of clusters.

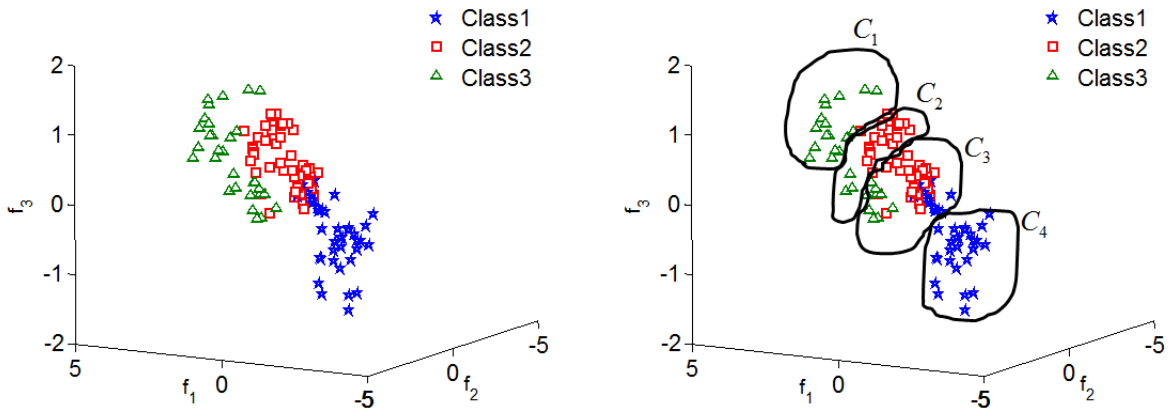
- Finally the probability of a signal belonging to a class is calculated according to the distribution of its representing feature vectors in different clusters and can be described as:

$$\Phi(e) = S \times M(:, e) \tag{7.13}$$

where $M(:, e)$ is the eth vector of the membership matrix, M, and the signal is labeled to belong to the class associated with the maximum value of $\Phi(e)$.

Example

The following example displays the proposed hard and fuzzy labeling in Fig. 7.6. The features



(a) (b)

Figure 7.6: (a) Three classes of a data set are shown in the feature domain. (b) Four clusters are identified in the feature space according to the relative structure of the feature samples in this plane.

of three synthetic classes are shown in Fig. 7.6(a) where each marking represents the features in each class. A clustering technique is applied to the features, and four clusters are estimated as shown in Fig. 7.6(b). The clusters are denoted with $\{C_1, C_2, C_3, C_4\}$. Table 7.1 displays the number of class 1 to 3 features in each cluster. To visualize the hard and fuzzy labeling methods

Table 7.1: The number of each features in each cluster (as performed in Fig. 7.6(b)).

	Class 1	Class 2	Class 3
C_1	0	0	16
C_2	0	20	3
C_3	15	30	9
C_4	35	0	0

proposed in this section, we use both methods to label the estimated clusters. According to the hard labeling, the class with the most member defines the label of a cluster. Therefore, using Table 7.1, $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ (the label of clusters 1 to 4) are $\{3, 2, 2, 1\}$, respectively. However, the fuzzy

technique calculates a membership matrix as explained in Eqns. 7.10 and 7.10:

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.87 & 0.13 \\ 0.28 & 0.55 & 0.17 \\ 1 & 0 & 0 \end{bmatrix} \tag{7.14}$$

Using the hard labeling, the second and third clusters were determined as class 2, while the fuzzy labeling assigned a relative membership coefficient to each class.

So far in this chapter, we introduced supervised cluster labeling with two different labeling approaches: hard and fuzzy labeling. In the rest of this chapter, we apply the developed techniques to two real-world applications. The first application which explains pathological voice classification is based on k-means clustering and hard labeling; and the second application uses SOTM and fuzzy labeling to classify environmental audio signals.

7.5 Experiment1: Pathological Voice Classification

Dysphonia or pathological voice refers to speech problems resulting from damage to or malformation of the speech organs. Dysphonia is more common in people who use their voice professionally; for example, teachers, lawyers, salespeople, actors and singers [143, 144], and it dramatically effects these professional groups's lives both financially and psychosocially [144]. In the past 20 years, a significant attention has been paid to the science of voice pathology diagnostic and monitoring. The purpose of this work is to help patients with pathological problems for monitoring their progress over the course of voice therapy. Currently, patients are required to routinely visit a specialist to follow up their progress. Moreover, the traditional ways to diagnose voice pathology are subjective, and depending on the experience of the specialist, different evaluations can be resulted. Developing an automated technique saves time for both the patients and the specialist, and can improve the accuracy of the assessments.

7.5.1 Background

Temporal features, such as, amplitude perturbation and pitch perturbation [145, 146] have been used for pathological speech classification; however, the temporal features alone are not enough for pathological voice analysis. Spectral and cepstral domains have also been used for pathological voice feature extraction; for example, mean fundamental frequency and standard deviation of the frequency [146], energy spectrum of a speech signal [147], mel-frequency cepstral coefficients (MFCC) [148] and linear prediction cepstral coefficients (LPCC) [149] have been used as pathological voice features. Gelzinis et al [150] and Senz-Lechna et al [151] provide a comprehensive review of the current pathological feature extraction methods and their outcomes. We mention only few of the techniques which reported a high accuracy; for example, Parsa in [152] achieves 96.5% classification using four fundamental frequency dependent features and two independent features based on the linear prediction (LP) modeling of vowel samples. In [149], Godino-Llorente et al feed MFCC coefficients of the vowel $/ah/$ from both normal and pathological speakers into a neural-network classifier, and achieve 96% classification rate. In [153], Umapathy et al present a new feature extraction methodology. In this paper, the authors propose a segment free approach to extract features such as octave max and mean, energy ratio and length and frequency ratio from the speech signals. This method was applied on continuous speech samples, and it resulted in 93.4% classification accuracy.

Based on the above, the majority of the current methods apply a short time spectrum analysis to the signal frames, and extract the spectral and temporal features from each frame. In other words, these methods assume the stationarity of the pathological speech over 10-30 ms intervals, and represent each frame with one feature vector; however, to our knowledge, the stationarity of the pathological speech over 10-30 ms has not been confirmed yet, and as a matter of fact, our observation from the TFD of abnormal speech evident that there are more transients in the abnormal signals, and the formants in pathological speech are more spread, and are less structured. Another shortcoming of the current approaches is that they require to segment the signal into short intervals. Using an appropriate signal segmentation has always been a controversial topic in windowed TF approaches. To overcome these limitations, we apply our proposed TFM feature

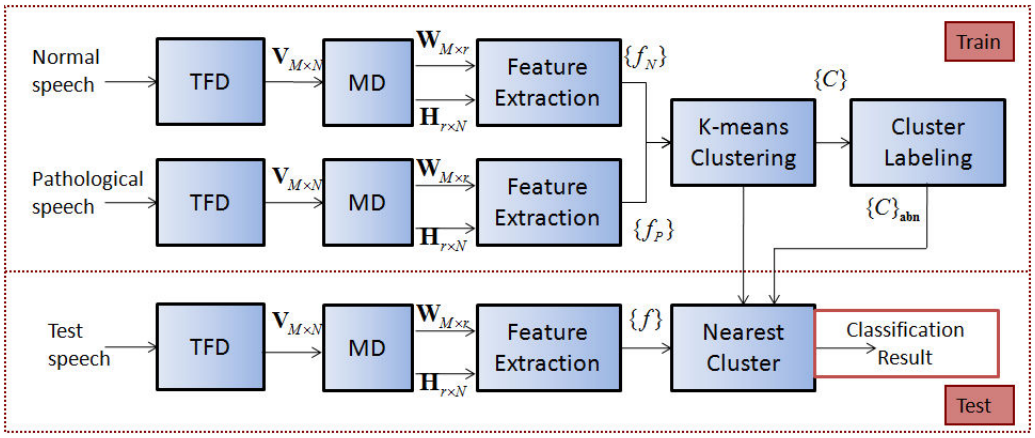


Figure 7.7: The schematic of the proposed pathological speech classification methodology.

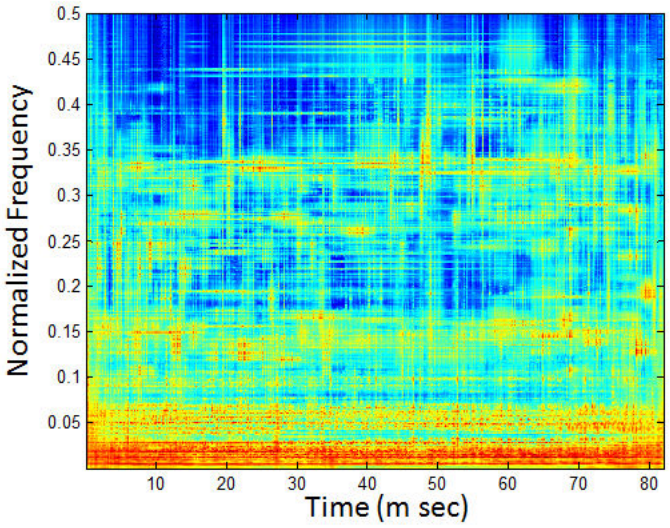
extraction technique to extract the TF features from the speech in a way that it captures the dynamic changes of the pathological speech, and apply our proposed supervised cluster labeling technique to classify the pathological speech signals.

7.5.2 Methodology

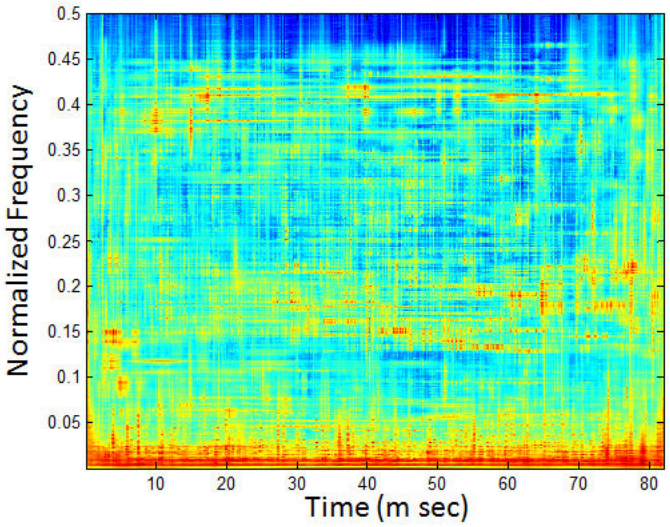
Fig. 7.7 is a schematic of the proposed pathological speech classification approach. As it can be seen in this figure, the system consists of two stages: learning and testing. In both stages, first the TFM decomposition is applied to the given signals. Next, TF features are extracted from each decomposed TF component.

Feature Extraction

In order to derive the discriminative features of normal and abnormal signals, we investigate the TFD difference of the two groups. To do so, we choose one normal and one pathological speech, and construct the Adaptive TFD 2.1 of each 80 ms frame of the signals. The sum of the TF matrices for each speech is shown in Fig. 7.8. We observed two major differences between the pathological and the normal speech: 1) the pathological signal has more transient components compared to the normal signal; and 2) the pathological voice presents weaker formants compared to the normal signal.



(a) TF distribution of a normal voice with a male speaker.



(b) TF distribution of a pathological voice with a male speaker.

Figure 7.8: TFD of a normal (a) and an abnormal signal (b) is constructed using adaptive TFD with Gabor atoms, 100 MP iterations and 5 MCE iterations. As evident in theses figures, the pathological signal has more transient components specially at high frequencies. In addition, the TF of the pathological signal presents weak formants, while the normal signal has more periodicity in low frequencies, and introduces stronger formants.

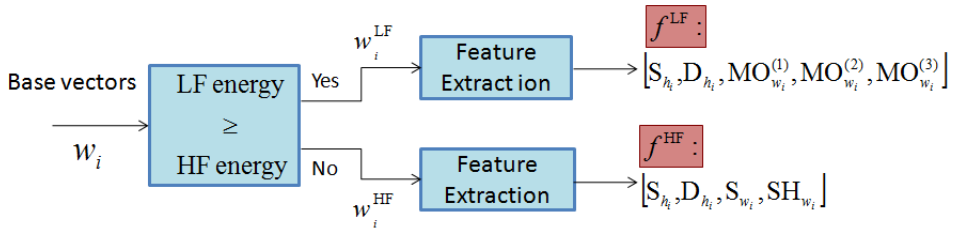


Figure 7.9: Block diagram of the proposed Feature extraction technique.

Additionally, our further observations of the decomposed components evidence that the abnormal speech behaves differently for voiced (vowel) and unvoiced (constant) components. Therefore, prior to extraction of the features from the decomposed components, we divide the base vectors into two groups: (i) *Low Frequency* (LF): the bases with dominant energy in the frequencies lower than 4 kHz, and (ii) *High Frequency* (HF): the bases with major energy concentration in the higher frequencies.

Based on the above observations, as depicted in Fig. 7.9, we extract four features from each LF-base and five features from each HF-base while only two of these two feature sets are the same. Except one of the features, Sharpness (SH_w), all the other features are extracted as explained in Section 5.3.

Sharpness: SH_w measures the spread of the components in low frequencies. In addition, we need another feature to provide an information on the energy distribution in frequency. Comparing the LF bases of the normal and the pathological signals, we notice that normal signals have strong formants; however, the pathological signals have weak and less structured formants. For each base vector, first we calculate the Fourier transform as given below:

$$W_i(\nu) = \left| \sum_{f=1}^{M} \mathrm{e}^{-j\frac{2\pi m\nu}{M}} w_i(m) \right| \tag{7.15}$$

where M is length of the base vector, and $W_i(\nu)$ is the Fourier transform of the base vector w_i. Next, we perform a second Fourier transform on the base vector, and obtain $W_i(\kappa)$ as the following:

$$W_i(\kappa) = \left| \sum_{\nu=1}^{M/2} \mathrm{e}^{-j\frac{2\pi\nu\kappa}{M/2}} W_i(\nu) \right| \tag{7.16}$$

Finally, we sum up all the values of $|W(\kappa)|$ for κ more than m_0, where m_0 is a small number:

$$\text{SH}_{w_i} = \sum_{\kappa=m_0}^{M/4} |W_i(\kappa)| \tag{7.17}$$

In order to demonstrate the behavior of feature SH_w, we assume that the base vector, w_i, has two components at frequencies samples m_1 and m_2 with energies of α and β respectively:

$$w_i(m) = \alpha\delta(m - m_1) + \beta\delta(m - m_2), \tag{7.18}$$

$|W(\nu)|$ (Eqn. 7.16) is calculated as below:

$$|W(\nu)| = \sqrt{\alpha^2 + \beta^2 + 2\alpha\beta\cos\left(2\pi\left(m_1 - m_2\right)\nu\right)} \tag{7.19}$$

$|W(\nu)|$ is independent to the parameter ν only when $m_1 \approx m_2$, or when the energy ratio of the components in Eqn. 7.18 is too small (either $\frac{\beta}{\alpha} \approx 0$ or $\frac{\alpha}{\beta} \approx 0$). In this case, when we calculate the Fourier transform of $|W(\nu)|$ as shown in Eqn. 7.16, $|W(\kappa)|$ is non-zero only at small values of κ (say $\kappa < m_0$, where m_0 is a small number). Hence, SH_{w_i} as it is calculated in Eqn. 7.17 results in a small feature. From the other side, $|W(\nu)|$ is dependent on the parameter ν when both the components in Eqn. 7.18 are strong ($\frac{\beta}{\alpha} \approx R, R \neq 0$). In this case, the Fourier transform of $|W(\nu)|$ is not negligible at $\kappa > m_0$, and SH_{w_i} results in lager values.

From the above explanation, we conclude that the small values of SH_{w_i} represent pathological formants, in which the components' energies are very small compared to the energy of the main frequency ($\frac{\beta}{\alpha} \approx 0$ or $\frac{\alpha}{\beta} \approx 0$), and the large values of SH_{w_i} shows the strong formants in speech ($\frac{\beta}{\alpha} \approx R, R \neq 0$).

Feature Clustering and Classification

As it is shown in Fig. 7.7, once the features are extracted, we feed them into a pattern classifier based on supervised cluster labeling explained earlier in this chapter. In the proposed technique, we use K-means clustering in a hard labeling approach.

Since separate features are extracted for LF and HF components, we have to train a separate classifier for each group: C^{LF} and C^{HF} for LF and HF components, respectively. Once the

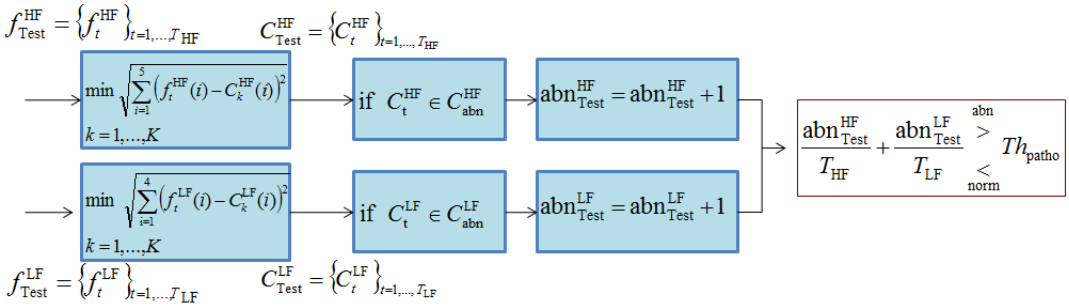


Figure 7.10: The block diagram of the test stage.

clusters are estimated, we count the number of abnormal feature vectors in each cluster, and the cluster with a majority of abnormal points is labeled as abnormal clusters; otherwise, the cluster is labeled as normal.

In the testing stage, we test the trained classifier. For a voice sample, we find the nearest cluster to each of its feature vectors using Euclidean distance criterion. If the number of the feature vectors that belong to the abnormality clusters is dominant, the voice sample is classified as a pathological voice; otherwise, it is classified as a normal speech.

Fig. 7.10 demonstrates the testing stage. $f^{\text{LF}}_{\text{Test}}$ and $f^{\text{HF}}_{\text{Test}}$ feature vectors are derived from the base and coefficient vectors in LF and HF groups, respectively. For each feature vector, we find the closest cluster, C_{u_0}, as given below:

$$\begin{aligned} f^{\text{LF}}_{t} \in C^{\text{LF}}_{u_0} \quad \text{if} \quad u_0 = \underset{k=1,...,K}{arg\,\min} \sqrt{\textstyle\sum_{i=1}^{4}\left(f^{\text{LF}}_t(i) - C^{\text{LF}}_k(i)\right)^2}, \\ {\scriptstyle t=1,...,T_{\text{LF}}} \\ f^{\text{HF}}_{t} \in C^{\text{HF}}_{u_0} \quad \text{if} \quad u_0 = \underset{k=1,...,K}{arg\,\min} \sqrt{\textstyle\sum_{i=1}^{5}\left(f^{\text{HF}}_t(i) - C^{\text{HF}}_k(i)\right)^2}, \\ {\scriptstyle t=1,...,T_{\text{HF}}} \end{aligned} \tag{7.20}$$

where, f^{LF}_t and f^{HF}_t are the input feature vectors, and T_{HF} and T_{LF} are the total numbers of test feature vectors for HF and LF components, respectively.

Next, the number of all the features that belong to abnormal and normal clusters is calculated:

$$\text{if } C^{\text{LF}}_{k_0} \in C^{\text{LF}}_{\text{abn}} \implies \text{abn}^{\text{LF}}_{\text{test}} = \text{abn}^{\text{LF}}_{\text{test}} + 1; \tag{7.21}$$

$$\text{if } C^{\text{HF}}_{k_0} \in C^{\text{HF}}_{\text{abn}} \implies \text{abn}^{\text{HF}}_{\text{test}} = \text{abn}^{\text{HF}}_{\text{test}} + 1; \tag{7.22}$$

where, $\text{abn}_{\text{test}}^{\text{LF}}$ and $\text{abn}_{\text{test}}^{\text{HF}}$ are the numbers of all the feature vectors of LF and HF groups that belong to an abnormal cluster. The signal is classified as normal if

$$L_{\text{abnormality}} < Th_{\text{patho}} \tag{7.23}$$

where Th_{patho} is the abnormality threshold, and $L_{\text{abnormality}}$ is the number of the abnormality features in the voice sample:

$$L_{\text{abnormality}} = \left(\frac{\text{abn}_{\text{test}}^{\text{LF}}}{T_{\text{LF}}} + \frac{\text{abn}_{\text{test}}^{\text{HF}}}{T_{\text{HF}}} \right) \tag{7.24}$$

If the criterion in Eqn. 7.23 is not satisfied, the signal is classified as a pathological speech.

7.5.3 Results and Discussions

The proposed methodology was applied to the Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database, distributed by Kay Elemetrics Corporation [84]. The database consists of 51 normal and 161 pathological speakers whose disorders spanned a variety of organic, neurological, traumatic, and psychogenic factors. The speech signal is sampled at 25 kHz and quantized at a resolution of 16 bits/sample. In this study, 25 abnormal and 25 normal signals were used to train the classifier.

MP-TFD with Gabor atoms is estimated for each 80 ms of the signal. Gabor atoms provide optimal TF resolution, and have been commonly used in MP-TFD. To acquire the required iterations (I) in the MP decomposition, we calculate the energy of the projected signal at each iteration, $\langle R^i x, g_{\gamma_i} \rangle$ in Eqn. 2.11. Fig. 7.11 illustrates the mean of the projected energy per iteration for one normal and one pathological signal. As evident in this figure, most of the coherent structure of the signal is projected before 100 iteration. Therefore, in this study, MP-TFD is constructed using the first 100 iterations and the remaining energy is ignored.

Next, we apply NMF with base number of $r = 15$ to each TF matrix, and estimate the base and coefficient matrices, **W** and **H** respectively. Each base vector is categorized into either LF or HF group: a base vector is grouped as LF component if its energy is concentrated more in the frequency range of 4 kHz or less; otherwise, it is grouped as HF component. We extract 4

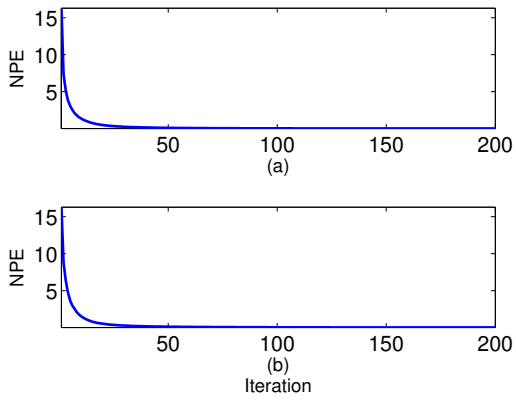


Figure 7.11: The Normalized projected energy (NPE) at each iteration is plotted for one normal (a) and one pathological signal (b). As it can be observed in this figure, most of the coherent structure of the signal is projected before 100 iteration, and the remaining energy is negligible.

features (S_h, D_h, S_w, SH_w) from each LF base vector w and its coefficient vector h, and 5 features ($S_h, D_h, MO_w^{(1)}, MO_w^{(2)}, MO_w^{(3)}$) from each HF base vector and its coefficient vector. In order to obtain the role of each feature in the classification accuracy, we calculate the p value of each feature using the Students t-test. The feature with the smallest p value plays the most important role in the classification accuracy. Fig. 7.12 demonstrates the relative importance of each 9 features. As shown in this figure, D_h and SH_w from LF features, and S_h, $MO_w^{(2)}$ and $MO_w^{(3)}$ from HF features play the most significant role in the classification accuracy.

Finally, we apply the K-means clustering to the logarithm of the derived feature vectors, and define the abnormality clusters. Figs. 7.13 illustrates the application of the proposed methodology for a pathological voice sample which is shown in Fig. 7.13(a). As explained earlier in this chapter, the test procedure determines the feature vectors that belong to the abnormality clusters. We use the base and coefficient matrices, $\mathbf{W}_{\text{abn}}$ and $\mathbf{H}_{\text{abn}}$, corresponding to the abnormality feature vectors, to reconstruct the abnormality TF matrix, $\mathbf{V}_{\text{abn}}$, as $\mathbf{V}_{\text{abn}} = \mathbf{W}_{\text{abn}}\mathbf{H}_{\text{abn}}$. Fig. 7.13(b) depicts the reconstructed TF matrix. As it is expected, the proposed method successfully distinguishes transients, high frequency components, and week formants as abnormality.

In the test stage, the trained classifier is used to calculate the measure of abnormality ($L_{\text{abnormality}}$ in Eqn. 7.24) for each voice sample. Fig. 7.14 shows the abnormality measure for 51 normal and

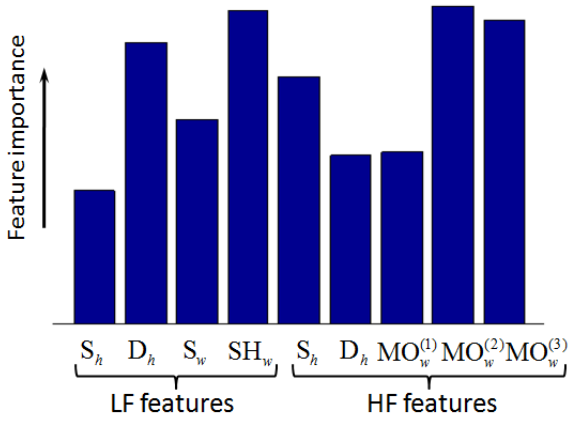


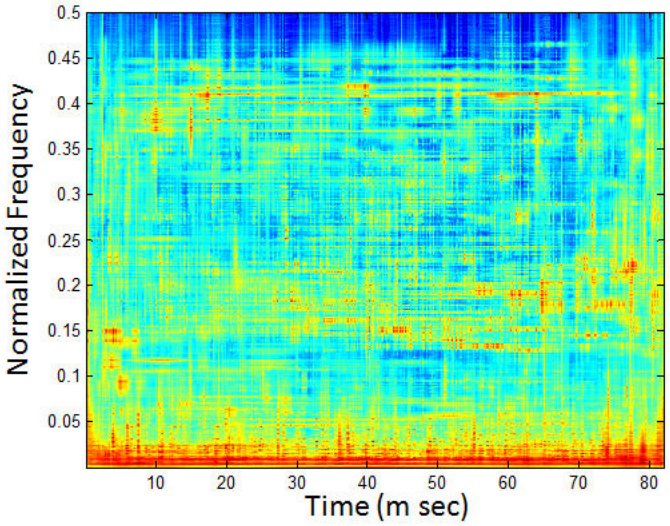
Figure 7.12: The relative height of each feature represents the relative importance of the feature compared to the other features.

161 pathological speech signals in MEEI database. As evident in this figure, the pathological samples have higher abnormality measure compared to the normal samples. Each signal is classified as normal if its abnormality measure is smaller than a threshold (Th_{patho} in Eqn. 7.23); otherwise it is classified as pathological. In order to find the abnormality threshold, receiver operating curves (ROC) of $L_{\text{abnormality}}$ is computed with the area under the curve indicating relative abnormality detection (Fig. 7.15). Based on the ROC, the cut-point of 0.59 is chosen as the abnormality threshold (Th_{patho} = 0.59). Table 7.2 shows the accuracy of the classifier. From the table, it can be

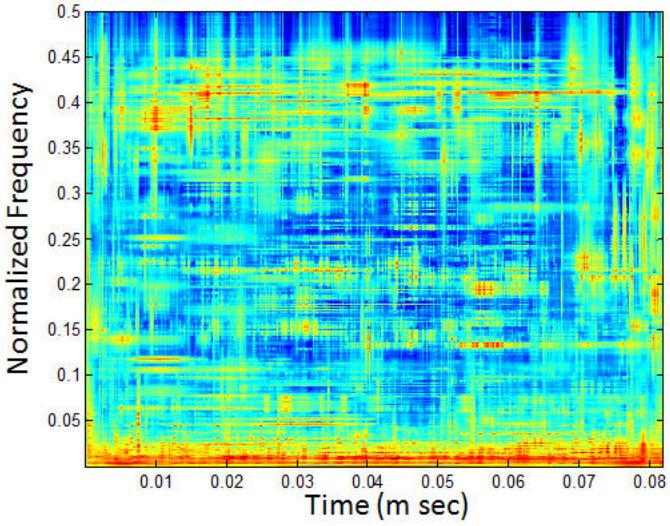
Table 7.2: Cluster labeling - Classification result.

Classes	Normal	Abnormal	Total
Normal	50	1	51
Pathological	2	159	161
Normal	98.0%	2.0%	100%
Pathological	1.2%	98.8%	100%

observed that out of 51 normal signals, 50 were classified as normal, and only 1 was misclassified as pathological. Also, the table shows that out of 161 pathological signals, 159 were classified as pathological and only 2 were misclassified as normal. The total classification accuracy is 98.6% which, to our knowledge, is the highest rate reported in literature. Additionally, in order to fur-



(a) TFD of a pathological speech.



(b) TFD of the estimated abnormality.

Figure 7.13: (a) TFM of a pathological speech signal. (b) The estimated abnormality TF matrix. As evident in this figure, the abnormality components are mainly transients, high frequency components, and weak formants.

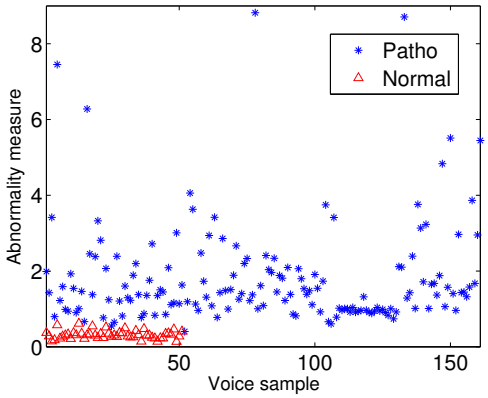


Figure 7.14: For each voice sample, the number of the feature vectors that belong to an abnormality cluster is calculated, and the abnormality measure is calculated as the ratio of the total number of the abnormal feature vectors to the total number of feature vectors in the voice sample.

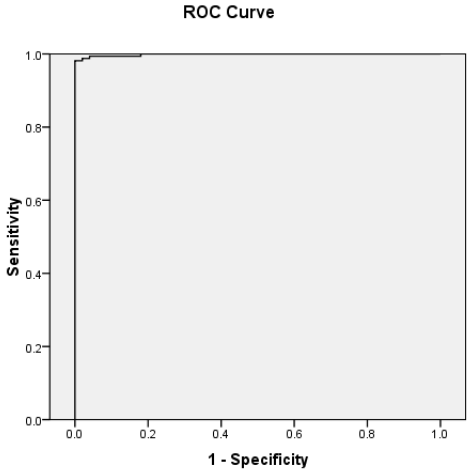


Figure 7.15: Receiver operating curve for the pathological voice classification is plotted. In this analysis, pathological speech is considered negative, and normal is considered positive. The area under the ROC is 0.999, and the maximum sensitivity for pathological speech detection while preserving 100% specificity is 98.1%.

ther compare the proposed cluster learning technique, we repeated the above experiment with a supervised learning method. In this approach instead of clustering techniques, we apply linear discriminant analysis (LDA) (explained in Section 6.6) as a well-known supervised classifier. The LDA technnique is applied to the extracted features, and the classification accuracy is reported in Table 7.3. As it can be seen in this table, the supervised method results in a very poor accuracy in case of pathological speech signals, while the cluster labeling technique provides a very high accuracy for both normal and pathological signals. The reason behind this significant improvement is the fact that the clustering method selects the abnormality features. These results taken together suggest that the proposed cluster labeling method successfully identifies the discriminant features of the abnormality speech signals.

Table 7.3: Supervised learning with LDA - Classification result.

Classes	Normal	Abnormal	Total
Normal	50	1	51
Pathological	51	110	161
Normal	98%	2%	100%
Pathological	32%	68%	100%

In Fig. 7.15 and Table 7.2, we utilized MD with decomposition order (r) of 15. We repeated the proposed method using different decomposition orders. Our experiment showed that the decomposition order of 5 and higher is suitable for our application. Table 7.4 shows the p values of three decomposition orders obtained with the Student's t-test.

Table 7.4: p value of the classifiers obtained with three different decomposition orders.

Decomposition order (r)	5	10	15
p value	3×10^{-10}	1×10^{-11}	1×10^{-13}

7.6 Experiment2: Environmental Audio Classification

The audio data set used in this work consists of 192 signals of about 3s duration, with a sampling rate of 22.05 kHz and a resolution of 16 bits per sample. A comprehensive information about dif-

ferent sound classes in the data set and the number of signals in each class was explained in Section 6.6. MP-TFD with frequency resolution of 44.1 Hz ($M = 250$) is constructed for each audio signal. Once the TFM ($\mathbf{V}_{250\times65533}$) is extracted, NMF with decomposition order of 15 ($r = 15$) is performed on each TFM. Next, nine features ($\mathrm{S}_{h_i}, \mathrm{D}_{h_i}, \mathrm{MO}_{w_i}, \mathrm{MO}_{h_i}, \mathrm{S}_{w_i}, \mathrm{SP}_{w_i}, \mathrm{P}_{w_i}, \mathrm{D}_{w_i}, \mathrm{MP}$) are extracted from each base and coefficient vector.

In classification stage, SOTM is applied on the training dataset and the number of valid clusters is calculated for each classification scenario. The clusters are formed, and the membership coefficients are calculated for each cluster based on the distribution of the train signals and the fuzzy labeling introduced in this chapter. In the test stage, each of the test signals are assigned to one of these cluster centers based on the minimum Euclidean distance measure. Finally, the class label of each signal is determined by the weighted sum of the feature vectors falling within each cluster multiplied by the membership coefficients. Another point to be discussed here is that since the data is represented to the SOTM in a random manner, the number and the shape and size of the clusters might vary each time the clustering algorithm is run on the data. However, since there is not a one to one correspondence between the clusters and the audio classes, this fact has no considerable impact on the total performance of the classifier.

7.6.1 Results and Discussions

One of the most important classification tasks for a hearing aid system is to discriminate human speech form environmental noise. Therefore, in the first scenario the data set consists of signals from human speech and environmental sounds. The human category includes 20 signals from male speakers and 20 signals from female speakers and environmental sounds include 10 bird, 10 aircraft, 10 piano and 10 animal signals. Table 7.5 shows the results for this classification task where an accuracy of 96% has been achieved. As it can be seen from the confusion matrices, the system demonstrates high accuracy in discrimination of human voice from other audio signals. The achieved true positive rate shows that all human voice signals have been classified correctly. In addition, the overall accuracy rate for classification scenarios that include discrimination of human voice is very high.

Table 7.5: Confusion matrix for classifying human vs non-human audio signals.

Classes	Human	Non-human	Total
Human	40 100%	0 0%	40 100%
Non-human	3 7.5%	37 92.5%	40 100%

Furthermore, in order to evaluate the efficiency of the system to discriminate human voice in particular environments, two other classification tasks have been defined. In the first case, an accuracy of 98% has been achieved in discrimination of human voice from the musical instruments. This capability could be useful in recognizing and separation of human voice from the background music in a song or at the concert. The second classification task was defined as discrimination of human voice from natural sounds, where an accuracy of 96% has been achieved. Furthermore, the proposed method was applied to other classification scenarios such as natural vs artificial sounds and musical instruments vs aircraft. Table 7.6 shows the overall obtained average accuracy rate and the data set used for each classification scenario.

Table 7.6: Different audio classes in the data set and the number of signals in each class.

Classes	Data Set	Average Accuracy
Human/Non-human	Non-human:aircraft, piano, animal, bird Human: male and female speeches	96%
Human/Music	Music: piano, flute, drum Human: male and female speeches	98%
Natural/Artificial	Natural: male, female, bird, animal, insect Artificial: helicopter, airplane, piano, flute, drum	81%
Human/Nature	Nature: animal, insect, bird Human: male and female speeches	98%
Aircraft/Music	Music: piano, flute, drum Aircraft: helicopter, airplane	98%

Additionally, we performed the classification over seven individual classifications, and reported the results using the fuzzy cluster labeling method and LDA-based supervised classification approach in Figs. 7.16 and 7.17, respectively. As it can be seen in the case of supervised learning, the

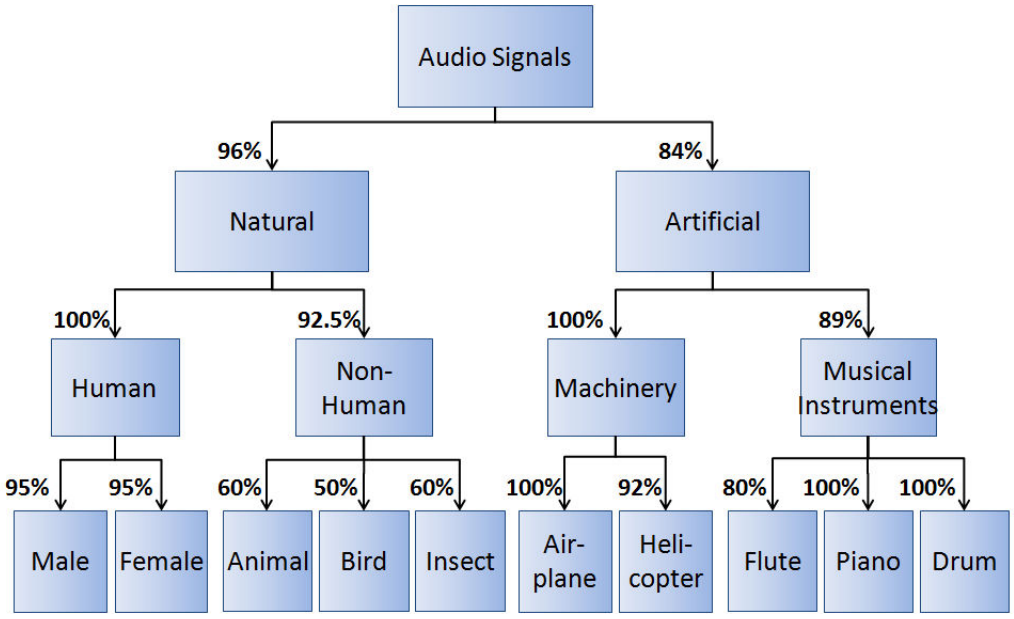


Figure 7.16: Fuzzy cluster labeling - Results for seven individual classifications over three levels.

accuracy is not significant which can be interpreted to be due to the overlapping features caused by the non-stationarity and diverse structures of the audio signals. These overlapping features do not let the supervised learning to find clearcut boarders between the feature points from different classes, which as a result limit the accuracy rate of the classification. Comparing Figs. 7.16 and 7.17, it can be seen that the cluster labeling method extensively improves the classification accuracy compared to the supervised learning. This approach adaptively assigns each feature vector to a class based on that classes membership, and it therefore increases the classification accuracy over all the three levels.

7.7 Chapter Summary

The contribution flowchart of this chapter is shown in Fig. 7.18. The objective was to improve the discrimination of the TF features in order to increase the performance of pattern recognition systems. To make this happen, this chapter presented a discriminant feature clustering technique based on a fusion of unsupervised and supervised learning approaches. This method applied an

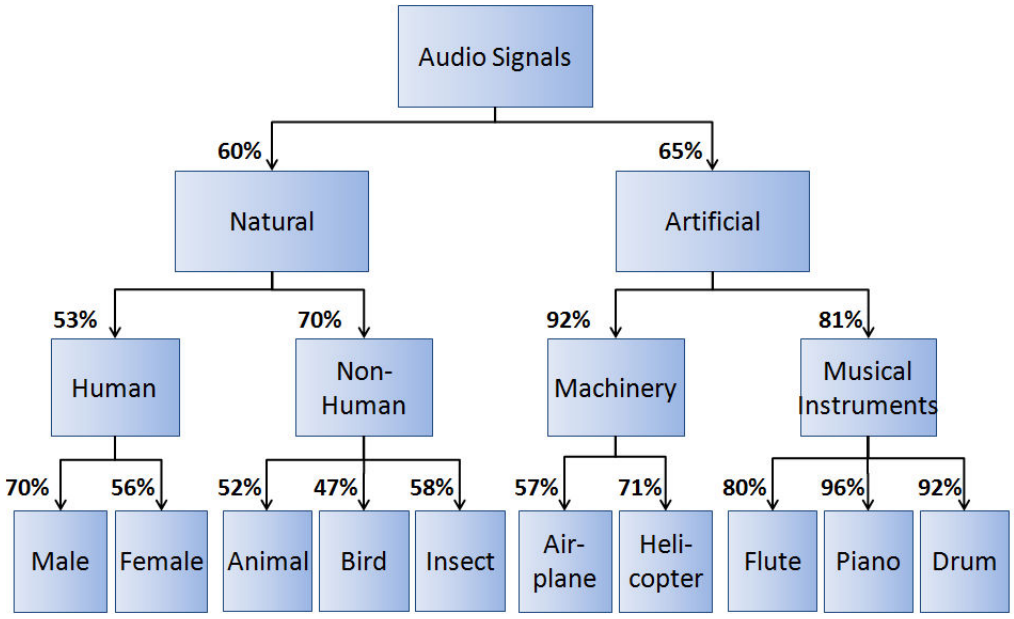


Figure 7.17: Supervised Learning with LDA - Results for seven individual classifications over three levels.

unsupervised clustering to all features, and then used a supervised labeling method to select the discriminative features. The supervised cluster labeling approach flexibly selected the feature points according to their importance for representing each corresponding class. The obtained features adaptively represented the complex discriminant patterns of real-world signals. Therefore, the selected features were better representatives of the real-world signals, and resulted in a significantly higher classification accuracy rate.

Furthermore, two cluster labeling methods were proposed: hard and fuzzy labellings. In hard labeling, each cluster was assigned to one of the possible classes, but in fuzzy labeling, each cluster was associated to each class with a relative membership value ranging from 0 to 1 (with 0 being the least contribution, and 1 being the most). Both proposed methods enhanced the commonly-used supervised learnings. In addition, we explained well-known clustering methods, and selected K-means and SOTM for the applications studied in this chapter. An advantage of SOTM compared to the K-means method was the number of clusters, which should be known beforehand in K-means, but was adaptively determined in the SOTM algorithm.

Although the proposed feature cluster labeling successfully identified discriminant features,

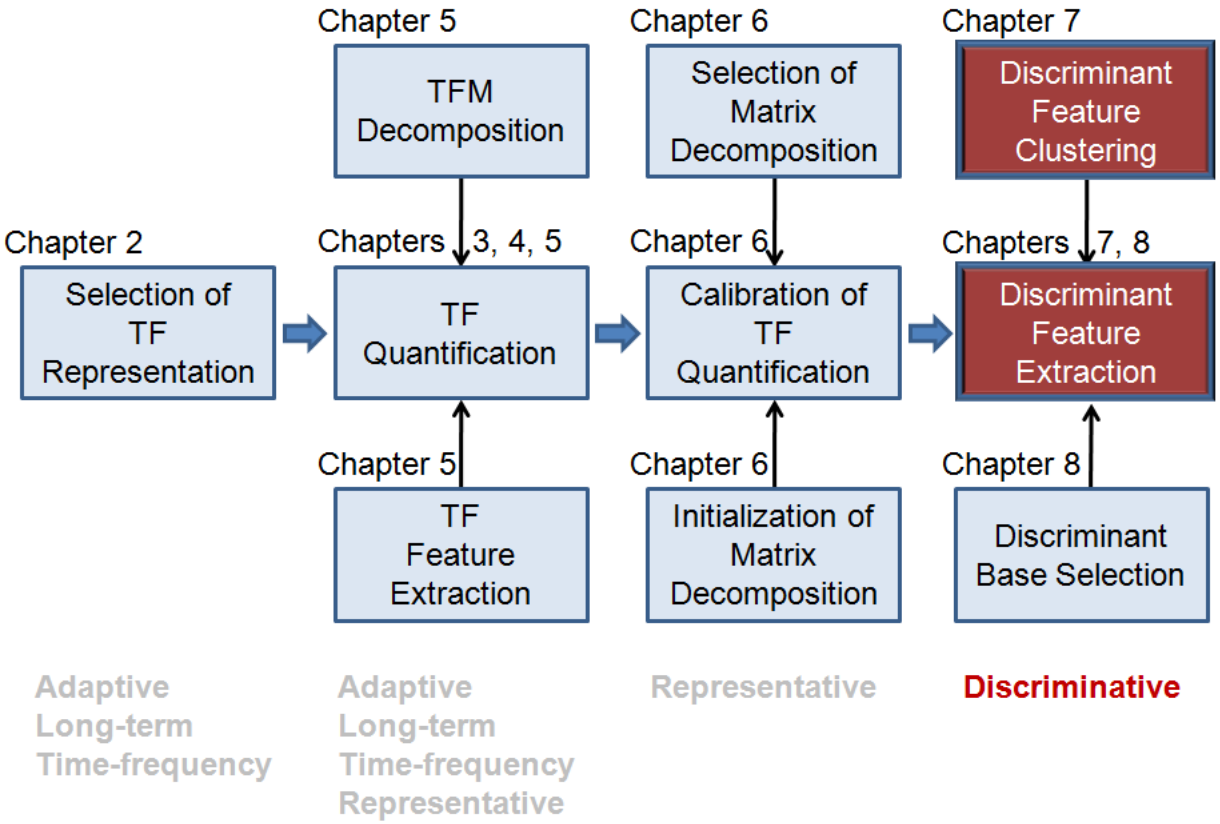


Figure 7.18: Flowchart of the proposed contributions.

some improvements can be made to the procedure to increase the efficiency of the approach. The method proposed in this chapter composed of two stages as it looks for the discriminant features after the features are extracted. In the next chapter, we create a new methodology that derives the discriminant features directly from the TFD. The proposed framework behaves as a fusion of the unsupervised feature clustering and supervised cluster labeling stages, and extracts the discriminant features in a very efficient manner.

Chapter 8

DISCRIMINANT BASES SELECTION IN TF MATRIX ANALYSIS

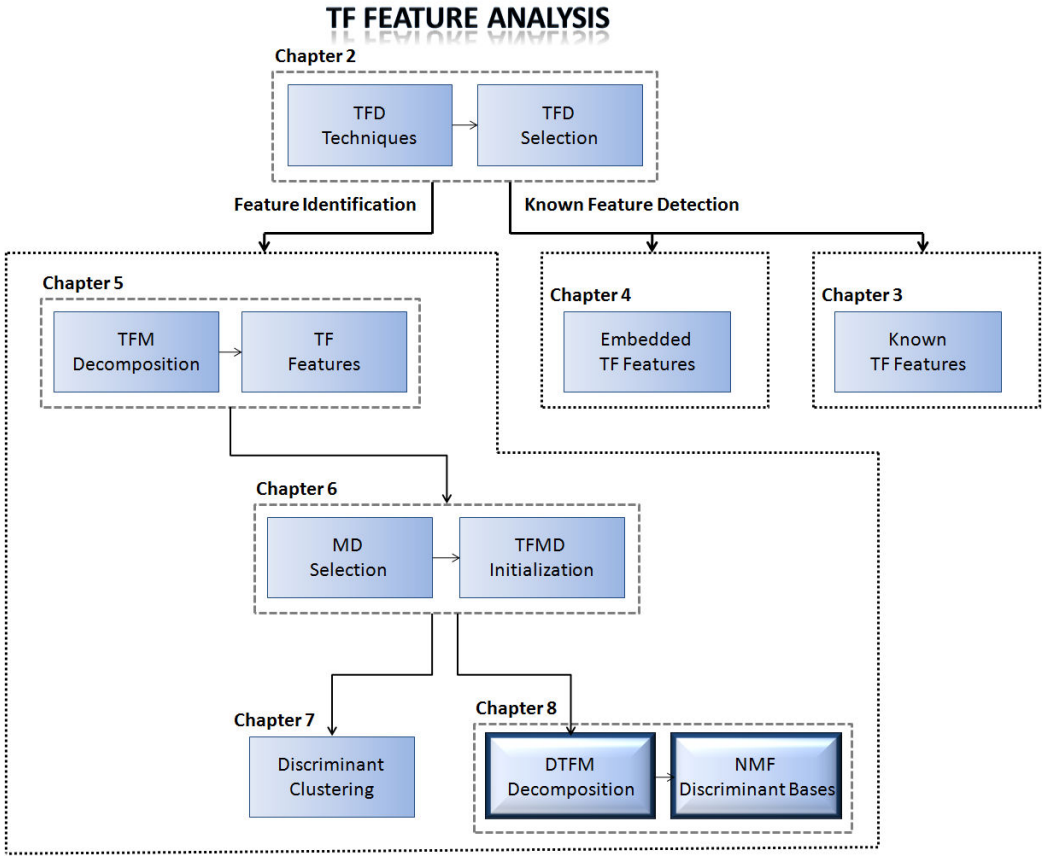


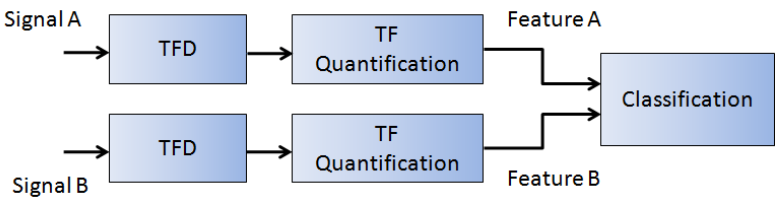
Figure 8.1: Chapter 8 - Selection of discriminant TF bases.

TIME or frequency descriptions, alone, are insufficient to provide comprehensive information about real world signals. On the contrary, time-frequency (TF) analysis is more suitable for revealing the non-stationary behavior of signals. The traditional TF approach assumes the stationarity of a signal over short segments and applies stationary tools to analyze each frame. However, the shortcoming of this classic approach is that first, it might segment the signals into parts that may not be considered stationary, and second, the approach does not use the long term information hidden in a signal. In this dissertation so far, we proposed TF matrix (TFM) quantification in which a matrix decomposition (MD) technique adaptively decomposes the TFM into components where the signal represents a similar TF behavior. Unlike the traditional methods, the TFM decomposition approach does not take any assumption about the stationarity of the signal. This approach adaptively defines the TF regions with the same spectral characteristics, and it therefore is a suitable tool for accurate representation and analysis of the time-varying behavior within non-stationary signals. In this chapter, our aim is to develop a new framework to identify discriminant TF bases that can be used for quantification and identification of the discriminant patterns in signals.

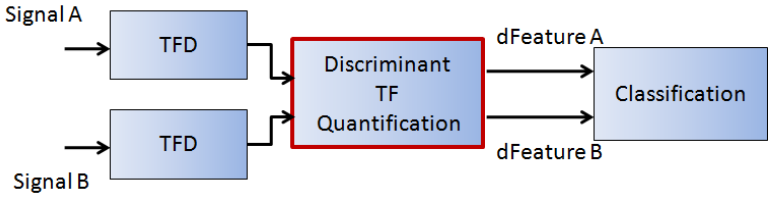
8.1 Motivation

Fig. 8.2(a) shows the general schematic of the current TF quantification methods in a two-group classification scheme. As it can seen in this figure, once the TFD is constructed, TF features are calculated from the TFD, and then a classifier is trained to discriminate the class A and B features in the feature space. As mentioned so far, a TF quantification notion is effective if the features derived from TF domain are not only representative of the TF structure, but also discriminative to the TF structures in the signals from different classes. In most of real world applications, the nature of signals from different classes are very similar, and there is only a few slight changes in the TF pattern of the signals that make the signals different from the signals of the other classes. Therefore, when extracting the TF features from different classes, the derived features might contain some overlapping as related to the similar TF structure.

In Chapter 5, we proposed a fusion of unsupervised clustering and supervised cluster label-



(a)



(b)

Figure 8.2: (a) The general block diagram of TF quantification. (b) The block diagram of proposed discriminant TF quantification.

ing technique to adaptively prioritize the clusters with discriminant features. This method was successfully applied to real-world signals, and showed a significant improvement over supervised classifiers. However, this approach will be more efficient if the discriminant features were extracted as part of the TF quantification stage rather as a post-processing tool. In Fig. 8.2(b), we display the schematic of such a discriminant TF quantification method. Instead of extracting each signals' feature set separately, we propose to quantify the TF distributions of the classes in such a way that the discriminant features are discriminated from the common ones. This method combines the TF quantification and the unsupervised clustering stages, and identifies the discriminant signatures of each class (denoted with 'dFeature A' and 'dFeature B' respectively), which then are used in the classification stage. The obtained discriminant features are expected to better characterize the differences between the signals, and as a result, improve the signal analysis performance in a variety of pattern recognition applications, such as signal classification and localization of region of discrimination (ROD).

Identifying the discriminant TF structure has attracted attention in the literature. Local discriminant base (LDB) analysis [22] is a wavelet packet based approach to identify the discriminative bases in the TF plane. While LBD analysis and its variants are an active area of research [23, 24, 25, 26], the optimal choice of LDBs highly depends on the nature of the dataset and the dissimilarity measures used to distinguish between classes. Also, LDB analysis can only be used with decomposition-based TF analysis such as wavelet. In this chapter, our aim is to propose a new discriminant analysis that is not restricted to any TF analysis approaches, and can be used along with any high time and frequency localized TF representation methods. The proposed framework, which we call discriminant TFM (DTFM) decomposition method, adaptively identifies the discriminant bases. This approach expands the TFM quantification concept and combines it with a new discriminant matrix decomposition (MD) method to adaptively identify the discriminant TF bases. To fulfill this objective, the present chapter aims to modify the TFM quantification approach, which proposed in Chapters 5 and 6, so that it estimates the discriminant bases. Once the discriminant bases are identified, they can be used to extract the key features of the discriminant patterns. The contributions of the present chapter are shown in Fig. 8.3.

8.2 Discriminant TFM Quantification

8.2.1 Methodology

In TFM quantification, we extract the features from the decomposed TF bases. If we distinguish the discriminative TF bases, then we can obtain more representative and discriminative features. Hence, in this chapter, our aim is to modify the TFM decomposition in a way that it identifies the discriminative TF bases between two classes, and call the modified technique discriminant TFM (DTFM) quantification.

As shown in Fig. 8.4, a two class (class A and class B) TFM quantification problem is considered. Our purpose is to extract the discriminating bases from two matrices $\mathbf{V}_1$ and $\mathbf{V}_2$, where $\mathbf{V}_1$ and $\mathbf{V}_2$ are the TFM of signals from class A and class B, respectively. The TFM quantification

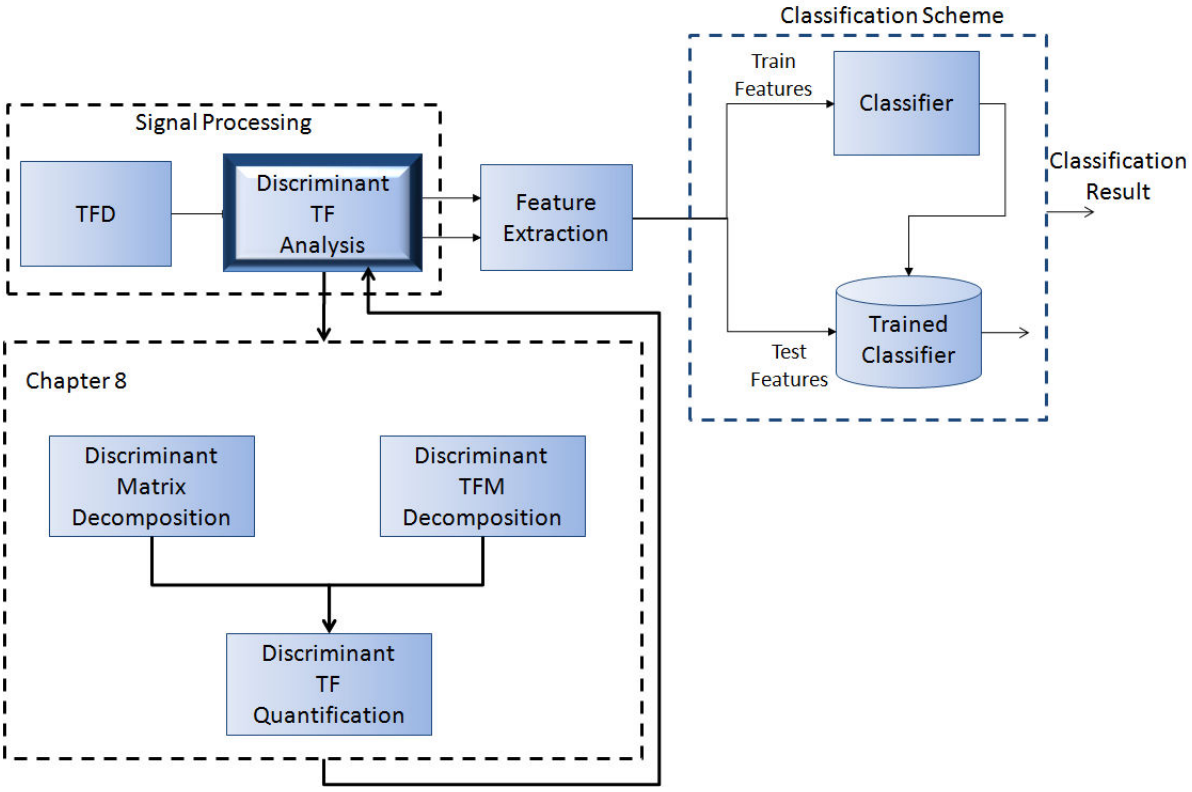


Figure 8.3: Chapter 8 - Discriminant Base Selection.

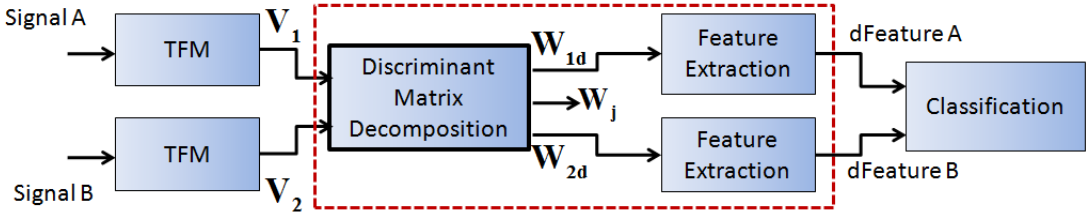


Figure 8.4: The block diagram of the discriminant TFM quantification approach.

decomposes each TF matrix into its components,

$$\mathbf{V}_1 = \mathbf{W}_1\mathbf{H}_1, \tag{8.1}$$
$$\mathbf{V}_2 = \mathbf{W}_2\mathbf{H}_2,$$

This approach decomposes the TFM of each class without any information from the TF structure of the other class, while the discriminant TFM quantification performs the decomposition considering the TF matrices from both of the classes. The new approach identifies the discriminant bases between the two classes as expressed in the following equations:

$$\mathbf{V}_1 = [\mathbf{W}_{1d} + \mathbf{W}_j]\,\mathbf{H}_{1d}, \tag{8.2}$$
$$\mathbf{V}_2 = [\mathbf{W}_{2d} + \mathbf{W}_j]\,\mathbf{H}_{2d}, \tag{8.3}$$

where $\mathbf{W}_{1d}$ and $\mathbf{W}_{2d}$ are the discriminant base matrices of class A and class B, respectively, and $\mathbf{W}_j$ contains the joint bases. The features calculated from $\mathbf{W}_{1d}$ and $\mathbf{W}_{2d}$ are expected to characterize the discriminant structure of each signal.

8.2.2 Visualization

The visualization of the proposed DTFM quantification is shown for the same synthetic example shown in Section 7.2 of Chapter 7. There are two signals in this example; one representing a normal class (class A) and the other one representing the abnormal class (class B). There are 6 components in each signal with three similar, and three different TF structures. The TFDs of these signals are shown in Figs. 8.5(a) and 8.5(b), respectively. The transients in the abnormal signal are the discriminant structure, and are used as abnormality descriptors.

Non-negative matrix factorization (NMF) is applied to decompose each TFM as explained in Eqn. 8.2. The decomposed bases ($\mathbf{W}_1$ and $\mathbf{W}_2$) are displayed in Figs. 8.5(c) and 8.5(d), respectively. Comparing the estimated base matrices, it can be observed that components 4, 3 and 1 from $\mathbf{W}_1$ are similar to the components 1, 5 and 4 in $\mathbf{W}_2$, respectively. These components represent the common structure in the TF structure of the two signals. It can also be seen that components 5, 6 and 2 from $\mathbf{W}_1$ and different from components 2, 3 and 6 in $\mathbf{W}_2$. This difference represents the discriminant structure that exist in one class but not in the other class.

Features are calculated from the decomposed matrices as shown in Fig. 8.5(e). The features corresponding to bases 5, 6 and 2 from class A are separated as 'dFeatureA', and the ones obtained from bases 2, 3 and 6 from class B are discriminated as 'dFeatureB'. Since the feature points in 'dFeatureA' and 'dFeatureB' are far from each other in the feature space, any simple classifier can find a boarder to separate these two modified feature sets and then successfully classify the normality and abnormality structures in the data. This example demonstrated that if the features are obtained from the discriminant bases, they represent the discriminant characteristics of each class.

To create a technique that adaptively distinguishes the discriminant bases of each signal, we develop a new discriminant TFM decomposition that will be explained in the next section.

8.3 Discriminant TFM Decomposition

In Chapter 4, we selected non-negative matrix factorization (NMF) as the desirable matrix decomposition approach as related to TF quantification. Over the last few years, several variants of NMF have been proposed. These methods impose additional constraints, such as localized or sparseness constraints, on the NMF cost function to construct a better NMF decompositions. Zafeiriou et al. [154] propose a discriminative NMF (DNMF), which adds an additional constraint seeking to maximize the between-class scatter and minimize the within-class scatter in the subspace spanned by the bases. However, none of the current methods has applied NMF for discriminant TFM quantification. In this chapter, we perform new constraints on NMF cost function to identify the discriminant bases, and denote the method as NMF Discriminant base (NMFDB) decomposition.

8.3.1 NMF Discriminant base (NMFDB)

In the two-class (class A and class B) TFM decomposition problem mentioned above, the traditional NMF decomposes each TFM into its components, $\mathbf{V}_1 = \mathbf{W}_1\mathbf{H}_1$ and $\mathbf{V}_2 = \mathbf{W}_2\mathbf{H}_2$, using the least square cost functions as follows:

$$D(\mathbf{V}_1, \mathbf{W}_1\mathbf{H}_1) = \|\mathbf{V}_1 - \mathbf{W}_1\mathbf{H}_1\|^2, \tag{8.4}$$

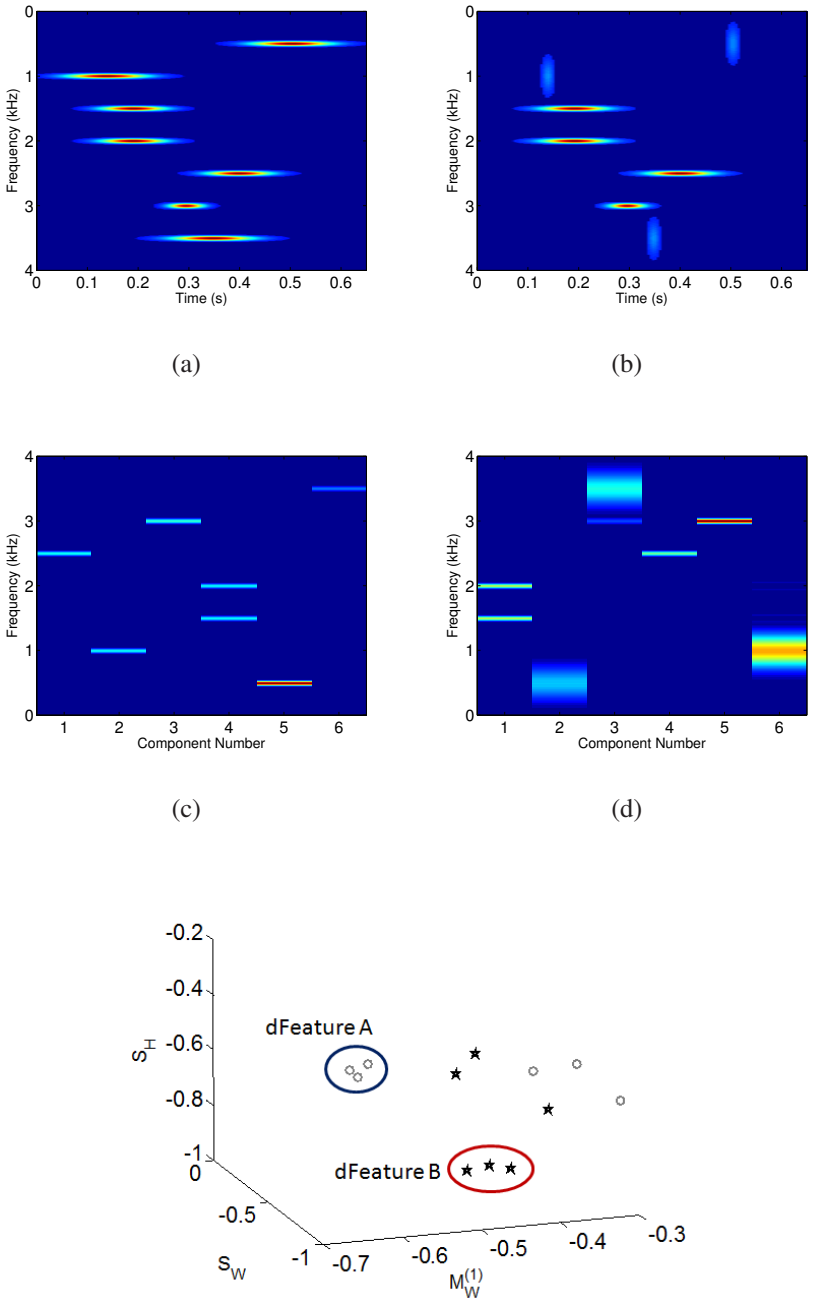


Figure 8.5: (a) TF representation of the normal signal. (b) TF distribution of the abnormal signal. (c) Decomposed spectral bases of the normal signal ($\mathbf{W}_1$)(d) Decomposed spectral bases of the abnormal signal ($\mathbf{W}_2$) (e) Feature space.

$$D(\mathbf{V}_2, \mathbf{W}_2\mathbf{H}_2) = \|\mathbf{V}_2 - \mathbf{W}_2\mathbf{H}_2\|^2, \tag{8.5}$$

where $\|.\|$ is the Euclidean norm. This approach decomposes the TFM of each class without any information from the TF structure of the other class. NMFDB performs the decomposition considering the TF matrices from both of the classes, and identifies the discriminant bases between the two classes as shown in Eqn. 8.3. The cost function of the NMFDB is modified as follows:

$$\begin{aligned} f = \quad & D(\mathbf{V}_1, [\mathbf{W}_{1d} + \mathbf{W}_j]\,\mathbf{H}_{1d}) + D(\mathbf{V}_2, [\mathbf{W}_{2d} + \mathbf{W}_j]\,\mathbf{H}_{2d}) \\ & -\delta D(\mathbf{W}_{1d}\mathbf{H}_{1d}, \mathbf{W}_2\mathbf{H}_{2d}) \\ & -\lambda D(\mathbf{W}_{2d}\mathbf{H}_{2d}, \mathbf{W}_j\mathbf{H}_{2d}) - \lambda D(\mathbf{W}_{1d}\mathbf{H}_{1d}, \mathbf{W}_j\mathbf{H}_{1d}), \end{aligned} \tag{8.6}$$

where δ and λ are positive constants. In Eqn. 8.6, the first two terms constrain the decomposed matrices to satisfy Eqn. 8.3, the second term maximizes the distance between discriminant base matrices, $\mathbf{W}_{1d}$ and $\mathbf{W}_{2d}$, and the last two terms maximize the distance between the discriminant base matrices and the joint base matrix. In this work, we adopted a projected gradient bound-constrained optimization method which is proposed by Lin [106].

8.3.2 Optimization of NMFDB

The optimization method is performed on the cost function f, and is consisted of five steps:

1) *Updating the Matrix* $\mathbf{W}_j$: In this stage, the optimization of the cost function f is solved with respect to $\mathbf{W}_j$, where f is the function given in Eqn. 8.6. In every iteration, matrix $\mathbf{W}_j$ is updated as below:

$$\mathbf{W}_j^{t+1} = \max\left\{\left(\mathbf{W}_j^t - \alpha^t \nabla f(\mathbf{W}_j^t)\right), 0\right\} \tag{8.7}$$

where t is the iteration order, $\nabla f(\mathbf{W}_j)$ is the projected gradient of the function f with respect to $\mathbf{W}_j$, while all the other matrices are constant, and α^t is the step size to update the matrix. $\nabla f(\mathbf{W}_j)$ is constructed from Eqn. 8.6 as given in the following equation:

$$\begin{aligned} \nabla f(\mathbf{W}_j) \quad = \quad & [((1+\lambda)\mathbf{W}_{1d} + (1-\lambda)\mathbf{W}_j)\,\mathbf{H}_{1d} - \mathbf{V}_1]\,\mathbf{H}_{1d}^{\mathrm{T}} \\ + \quad & [((1+\lambda)\mathbf{W}_{2d} + (1-\lambda)\mathbf{W}_j)\,\mathbf{H}_{2d} - \mathbf{V}_2]\,\mathbf{H}_{2d}^{\mathrm{T}}, \end{aligned} \tag{8.8}$$

The step size is found as $\alpha^t = \beta^{K_t}$. Where $\beta^1, \beta^2, \beta^3, \ldots$ are the possible step sizes, and K_t is the first non-negative integer for which:

$$f(\mathbf{W}_j^{t+1}) - f(\mathbf{W}_j^t) \leq \sigma \left\langle \nabla f(\mathbf{W}_j^t), \mathbf{W}_j^{t+1} - \mathbf{W}_j^t \right\rangle \tag{8.9}$$

where the operator $\langle ., . \rangle$ is the inner product between two matrices. In [106], values of σ and β are suggested to be 0.01 and 0.1, respectively. Once the step size, α^t, is found, the stationarity condition of function $f(\mathbf{W}_j)$ at the updated matrix is checked as below:

$$\left\| \nabla^P f(\mathbf{W}_j^{t+1}) \right\| \leq \epsilon \left\| \nabla f(\mathbf{W}_j^1) \right\| \tag{8.10}$$

where $\left\| \nabla f(\mathbf{W}_j^1) \right\|$ is the the projected gradient of the function $f(\mathbf{W}_j)$ at first iteration ($t = 1$), ϵ is a very small tolerance, and $\nabla^P f(\mathbf{W}_j)$ is the projected gradient defined as:

$$\nabla^P f(\mathbf{W}_j) = \begin{cases} \nabla f(\mathbf{W}_j), & w_{mr} > 0, \\ \min\left(0, \nabla f(\mathbf{W}_j)\right), & w_{mr} = 0, \end{cases} \tag{8.11}$$

If the stationary condition is met, the procedure stops, if not, the optimization is repeated until the point $\mathbf{W}_j^{t+1}$ becomes a stationary point of $f(\mathbf{W}_j)$.

2,3,4,5) *Updating the Matrices* $\mathbf{W}_{1d}$, $\mathbf{W}_{2d}$, $\mathbf{H}_{1d}$ and $\mathbf{H}_{2d}$: These stages solve the optimization problem respect to $\mathbf{W}_{1d}$, $\mathbf{W}_{2d}$, $\mathbf{H}_{1d}$ and $\mathbf{H}_{2d}$ assuming the remaining matrices are constant. A similar procedure to what we did in stage 1 is repeated in here. Calculation of $\nabla f(\mathbf{W}_{1d})$, $\nabla f(\mathbf{W}_{2d})$, $\nabla f(\mathbf{H}_{1d})$ and $\nabla f(\mathbf{H}_{2d})$ are calculated as follows:

$$\begin{aligned}
\nabla f(\mathbf{W}_{1d}) &= [(1-\delta-\lambda)\mathbf{W}_{1d} + (1+\lambda)\mathbf{W}_j \\
&+ \mathbf{W}_{2d}\mathbf{H}_{2d} - \mathbf{V}_1]\,\mathbf{H}_{1d}^{\mathrm{T}}, \qquad (8.12) \\
\nabla f(\mathbf{H}_{1d}) &= \mathbf{W}_{1d}^{\mathrm{T}}\left[((1-\delta-\lambda)\mathbf{W}_{1d} + (1+\lambda)\mathbf{W}_j)\right]\mathbf{H}_{1d} \\
&+ \mathbf{W}_j^{\mathrm{T}}\left[(1+\lambda)\mathbf{W}_{1d} + (1-\lambda)\mathbf{W}_j\right]\mathbf{H}_{1d} \\
&+ \delta\mathbf{W}_{1d}^{\mathrm{T}}\mathbf{W}_{2d}\mathbf{H}_{2d} - \left(\mathbf{W}_{1d} + \mathbf{W}_{2d}\right)^{\mathrm{T}}\mathbf{V}_1, \qquad (8.13) \\
\nabla f(\mathbf{W}_{2d}) &= [(1-\delta-\lambda)\mathbf{W}_{2d} + (1+\lambda)\mathbf{W}_j \\
&+ \mathbf{W}_{1d}\mathbf{H}_{1d} - \mathbf{V}_2]\,\mathbf{H}_{2d}^{\mathrm{T}}, \qquad (8.14) \\
\nabla f(\mathbf{H}_{2d}) &= \mathbf{W}_{2d}^{\mathrm{T}}\left[((1-\delta-\lambda)\mathbf{W}_{2d} + (1+\lambda)\mathbf{W}_j)\right]\mathbf{H}_{2d}
\end{aligned}$$

$$+ \quad \mathbf{W}_j^{\mathrm{T}} \left[(1+\lambda)\mathbf{W}_{2d} + (1-\lambda)\mathbf{W}_j\right] \mathbf{H}_{2d}$$
$$+ \quad \delta \mathbf{W}_{2d}^{\mathrm{T}} \mathbf{W}_{1d} \mathbf{H}_{1d} - \left(\mathbf{W}_{2d} + \mathbf{W}_{1d}\right)^{\mathrm{T}} \mathbf{V}_2, \tag{8.15}$$

The steps 1 to 5 are iteratively repeated. The optimization process is complete when the cost function is smaller than a threshold, or a certain number of iterations is reached.

8.3.3 Visualization of NMFDB

The performance of NMFDB is visualized through a synthetic example. In this example two signals are created: Signal A and Signal B. Signal A, y_1, is chosen to be the signal generated according to the below equation:

$$x(t) = \sum_{i=1}^{7} x_i(t) = \sum_{i=1}^{7} \alpha_i g\left(\sigma_i, \mu_i\right) \sin\left(a_i t\right), \tag{8.16}$$

where $g\left(\sigma, \mu\right)$ is a Gaussian with mean μ and variance of σ^2. (α, σ, μ, a) for each component from 1 to 7 is as following: (3,0.001,0.42,2π 3600), (1,0.05,0.68,2π 2600), (1,0.05,0.68,2π 600), (3,0.008,0.8,2π 1700), (3,0.008,1.27,2π 1700), (1,0.04,0.93,2π 1000), (1,0.03,1.18,2π 2600). Signal B, y_2, is created from Signal A by removing one of the TF functions (g_{γ_5}). Fig. 8.6 shows the time and TFM plots of A and B signals.

We applied NMFDB to the TF matrices of y_1 and y_2. In this example, it is expected that NMFDB identifies the difference in the TF structure of the two signals. Figs. 8.7(a) (b) and (c) show the decomposition matrices, $\mathbf{W}_j$, $\mathbf{W}_1$ and $\mathbf{W}_2$, respectively. $\mathbf{W}_j$ shows the similar bases in the signals A and B, and $\mathbf{W}_1$ and $\mathbf{W}_2$ represent the discriminant TF bases in each signal.

The decomposed matrices are used to construct the joint and the discriminant matrices as $\mathbf{V}_{j1} = \mathbf{W}_j\mathbf{H}_1$, $\mathbf{V}_{j2} = \mathbf{W}_j\mathbf{H}_2$, $\mathbf{V}_{d1} = \mathbf{W}_1\mathbf{H}_1$ and $\mathbf{V}_{d2} = \mathbf{W}_2\mathbf{H}_2$, where $\mathbf{V}_{j1}$ and $\mathbf{V}_{j2}$ are the similar TF structures in signal A and signal B, respectively, $\mathbf{V}_{d1}$ is the discriminant TF structure in signal A, and $\mathbf{V}_{d2}$ is the discriminant TF structure in signal B. The constructed matrices are shown in Fig. 8.7. As evident in Figs. 8.7(d) and (e), all the components except g_{γ_5} are successfully captured as the similar TF structure of the two signals. Fig. 8.7(f) shows the discriminant TF structure in signal A ($\mathbf{V}_{d1}$). As expected, g_{γ_5} is identified as the discriminant function in signal A, as shown in

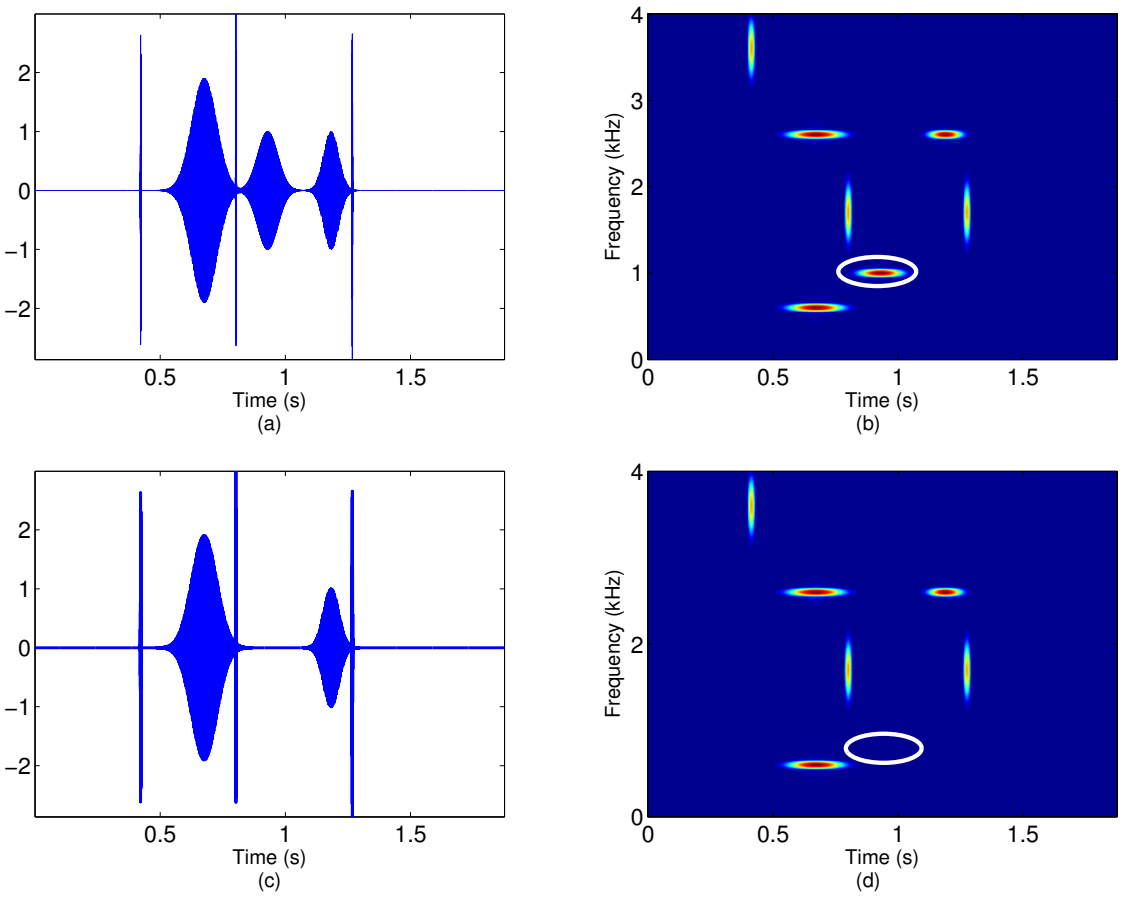


Figure 8.6: Two synthetic signals A and B are generated. The time and spectrogram plots of signal A are shown in Figs. (a) and (b), respectively, and Figs. (c) and (d) belong to the time and spectrogram plot of signal B, respectively. The ellipse shows the location of the TF difference in the signals A and B.

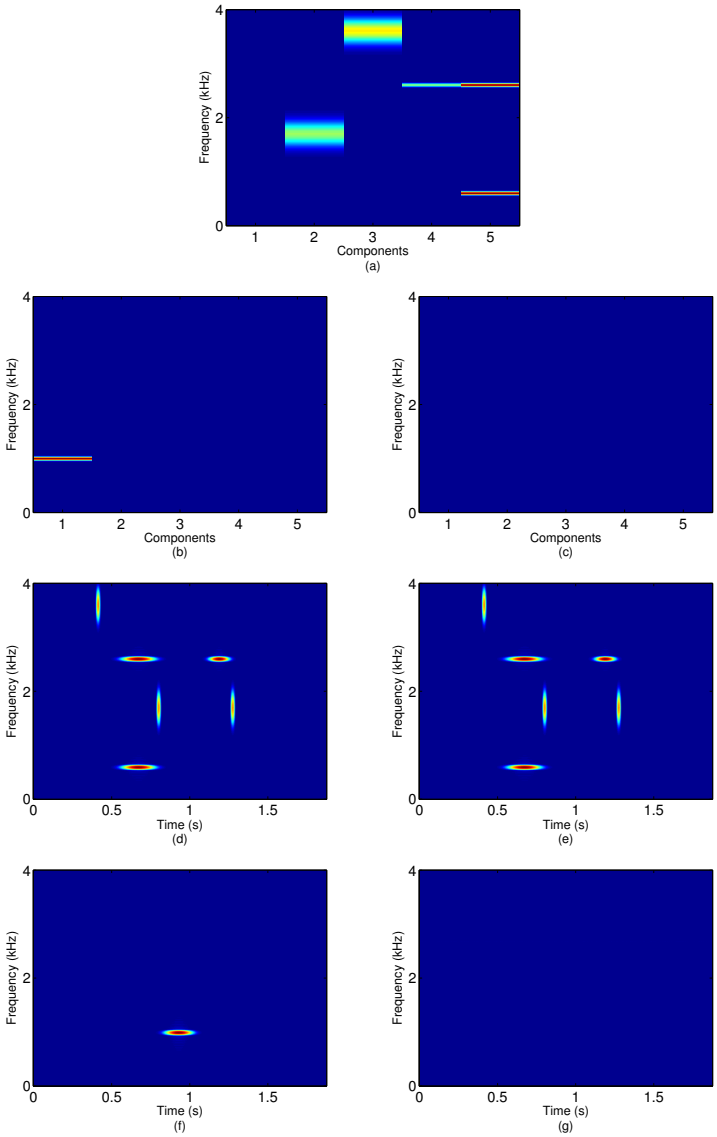


Figure 8.7: Visualization of the NMFDB method. (a) The common bases between signals A and B ($\mathbf{W}_j$). (b) The discriminant bases of signal A ($\mathbf{W}_1$). (c) The discriminant bases of signal B ($\mathbf{W}_2$). (d) The common TF structure in signal A. (e) The common TF structure in signal B. (f) The discriminant TF structure in signal A. (g) The discriminant TF structure in signal B. As expected, g_{γ_5} is identified as the TF difference in signal A, and no discriminant TF structure is identified in signal B.

Fig. 8.7(g), the discriminant matrix of signal B is all zero, which means that all the functions in signal B are also present in signal A.

The example demonstrated above showed that the NMFDB algorithm successfully identified the discriminant TF bases in signal A. In many real world applications, a signal may undergo operations such as amplitude scaling or temporal shift. For example, in speaker recognition application, the speech signals recorded from a speaker could have different amplitudes, or there could be a time difference in the recorded voices. From pattern recognition point of view, it is important that the identified discriminant bases are insensitive to these signal variations, and are able to retrieve the patterns even after these operations. In the subsequent section, we present the analysis of NMFDB properties with respect to two most common operations.

8.4 Properties of DTFM Decomposition

It is desirable that the extracted discriminant bases are robust to signal processing operations. This section examines the properties of NMFDB with respect to amplitude scaling and time shift which are the most common signal processing operations.

8.4.1 Amplitude Scaling

If $y(t)$ be the amplitude scaled version of signal $x(t)$, such that:

$$y(t) = ax(t), \tag{8.17}$$

the TFM of the signal $y(t)$, $\mathbf{V}_y$, can be written as follows:

$$\mathbf{V}_y = a^2\mathbf{V}_x, \tag{8.18}$$

where $\mathbf{V}_x$ is the TFM of signal $x(t)$ and a is the amplitude scale. Fig. 8.8 shows $x(t)$ and $y(t)$ signals for $a = 2$. The amplitude scaling does not effect the TF structure of the signal, therefore, NMFDB should not identify any discriminant bases. We applied the NMFDB to the TF matrices of the two signals, and as it is expected, the discriminant matrices, $\mathbf{W}_1$ and $\mathbf{W}_2$, both were empty.

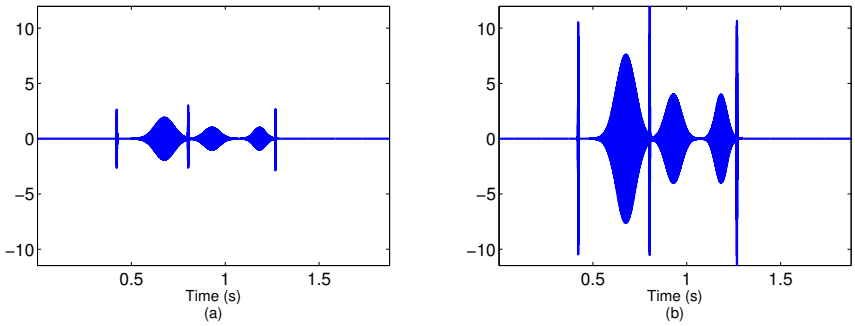


Figure 8.8: NMFDB amplitude scaling property. (a) Signal $x(t)$. (b) Signal $y(t) = 4x(t)$. NMFDB identified no discriminant TF structure between these two signals.

8.4.2 Time Shift

The property of NMFDB under temporal shift is examined using two examples. Let us call the signal in Figure 8.9 (a) with $x(t)$ and its corresponding TFM with $\mathbf{V}_x$ (shown in Figure 8.9 (b)). We perform a circular shift of τ ms to the signal $x(t)$ as follows:

$$y(t) = x(t - \tau), \tag{8.19}$$

The temporal shift only shifts the TF structure of the TFM in time, and does not change the TF components in the signal. Hence, the TFM of the signal $y(t)$, $\mathbf{V}_y$ can be written as follows:

$$\mathbf{V}_y(f, t) = \mathbf{V}_x(f, t - \tau), \tag{8.20}$$

The temporal signal and its corresponding TFM under the time shift of 250 ms (τ = 250 ms) are shown in Figs. 8.9 (c) and (h), respectively. NMFDB decomposed the TF matrices, $\mathbf{V}_y$ and $\mathbf{V}_x$, as follows:

$$\mathbf{V}_x = [\mathbf{W}_x + \mathbf{W}_j]\,\mathbf{H}_x, \tag{8.21}$$

$$\mathbf{V}_y = [\mathbf{W}_y + \mathbf{W}_j]\,\mathbf{H}_y, \tag{8.22}$$

NMFDB did not identify any discriminant TF bases; the decomposed discriminant matrices, $\mathbf{W}_x$ and $\mathbf{W}_y$, were found empty. $\mathbf{W}_j$ contained the spectral vectors that are common in $\mathbf{V}_x$ and $\mathbf{V}_y$, and $\mathbf{H}_y$ was equal to $\mathbf{H}_x(t - \tau)$. We repeated the time shift example with circular shift of τ = 750 ms.

The shifted signal and its TFM can be seen in Figs. 8.9 (d) and (i). Similar to the above example, NMFDB did not identify any discriminant TF bases. The joint bases ($\mathbf{W}_j$) and the corresponding coefficient matrices ($\mathbf{H}_x$ and $\mathbf{H}_y$) are shown in Fig. 8.9. As evident in this figure, for both time shift values, $\mathbf{H}_y$ is found to be equal to $\mathbf{H}_x(t-\tau)$.

8.5 Experiment: Synthetic Signal and Pathological Speech

The discriminant TFM (DTFM) decomposition method could have important implications in signal analysis. Three main applications include the detection of the discriminant bases, localization of the region of discrimination (ROD) and signal classification that are shown in this section.

8.5.1 Detection of the discriminant bases

In this example, we create two signals with a known discriminant pattern, and examine the capability of DTFM to identify the bases that represent the existent discriminant pattern. First, two signals are constructed: signal A and signal B, where signal A is a piano signal, y_1, of 350 ms with sampling frequency of 44 kHz, and signal B is the same piano signal added to a linear chirp as denoted below:

$$\begin{aligned} y_2(t) &= y_1(t) + y_c(t), \\ &= y_1(t) + \kappa \cos\left\{2\pi\left(f_0 t + c t^2\right)\right\} G(\mu, \sigma), \end{aligned}$$

where $c = (f_1 - f_0)/t_s$, f_0 is the initial frequency at $t = 0$, f_1 is the instantaneous frequency at $t = t_s$ and κ is the weight scale. $G(\mu, \sigma)$ is a Gaussian function with mean and variance of μ and σ, respectively that defines the location and the duration of the chirp signal. In this experiment, $f_0 = 20$ Hz, $f_1 = 5.5$ kHz, $t_s = 0.09$ s and $(\mu, \sigma) = (1.4, 0.01)$. Fig. 8.10 shows y_1 and y_2 signals and their corresponding TF matrices, $\mathbf{V}_1$ and $\mathbf{V}_2$ when $\kappa = 0.2$.

Second, the NMFDB decomposition is applied to the TF matrices of A and B signals as in Eqn. 8.3. The obtained joint base matrix ($\mathbf{W}_j$), the discriminant base matrices ($\mathbf{W}_1$ and $\mathbf{W}_2$) and the corresponding coefficient vectors are used to estimate the common and the discriminative TF structures as shown in Fig. 8.10. As evident in this figure, DTFM correctly identified the spectral characteristic of the chirp as the discriminant pattern.

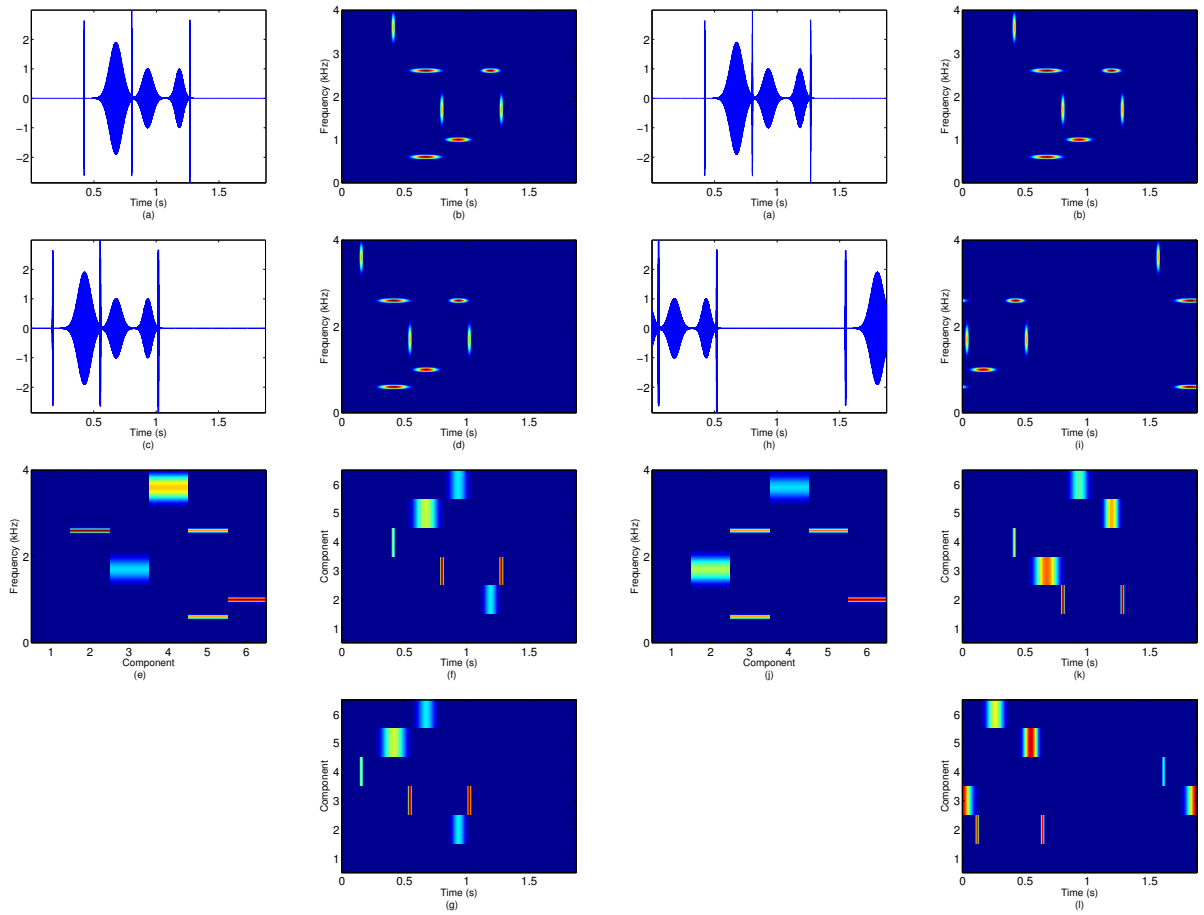


Figure 8.9: NMFDB temporal shift property. (c) The signal with 250 ms temporal shift to the left. (d) TFM of the signal in (c). (e) The common bases ($\mathbf{W}_j$). (f) The coefficient matrix of the original signal, $\mathbf{H}_x$. (g) The coefficient matrix of the shifted signal, $\mathbf{H}_y$. (h) The signal with 750 ms circular shift to the left. (i) TFM of the signal in (h). (j) The common bases ($\mathbf{W}_j$). (k) The coefficient matrix of the original signal, $\mathbf{H}_x$. (l) The coefficient matrix of the shifted signal, $\mathbf{H}_y$.

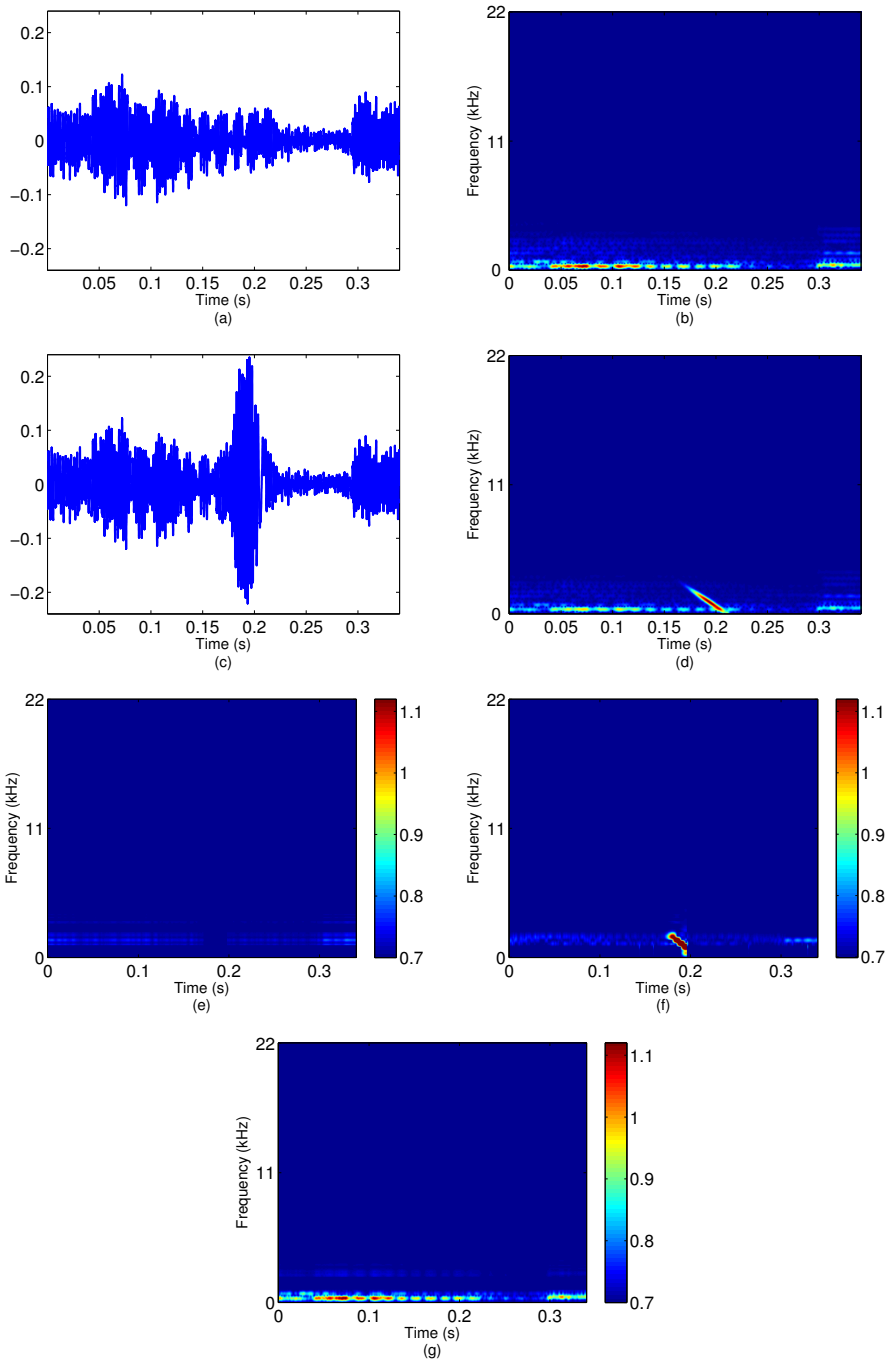


Figure 8.10: The DTFM method detects the discriminant pattern in two signals. (a) A 350 ms segment of a piano signal ($y_1(t)$). (b) The TFM of the piano segment. (c) A signal ($y_2(t)$) is generated by adding a chirp with $\kappa = 0.2$ to the piano segment as identified in Eqn. 8.23. (d) The TFM of the piano + chirp segment. (e) The discriminant TF pattern detected in the piano signal, $y_1(t)$. (f) The discriminant TF pattern identified in $y_2(t)$ signal. (g) The common TF structure detect.

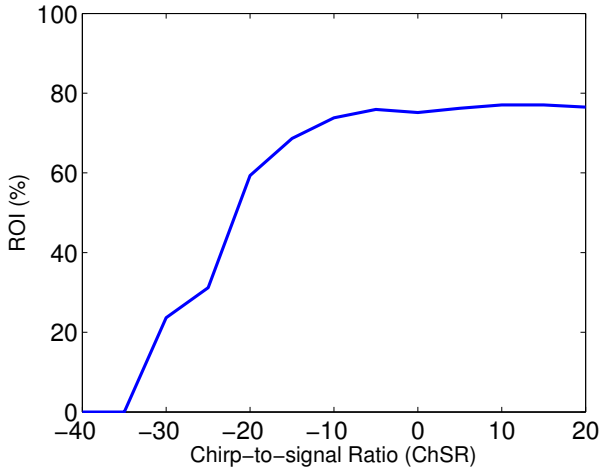


Figure 8.11: The localization percentage of NMFDB when ChSR decreases from 20dB to -40dB. It can be seen that at ChSR of -30 db, NMFDB still localizes 20% of the difference.

Finally, we demonstrate the robustness of DTFM as the energy of the discriminative pattern decreases. In Fig. 8.11, ROD(%) shows the accuracy of the DTFM method to localize the discriminative pattern as per chirp-to-signal ratio (ChSR). ChSR represents the amount of chirp signal that is added to the signal as defined below:

$$ChSR(db) = 10\log \sum y_c^2 - 10\log \sum y_1^2, \tag{8.23}$$

and ROD(%) is quantified as the correlation between the identified TFM and the actual discriminant TFM as follows:

$$ROD(\%) = \frac{100 \sum_m \sum_n (\mathbf{V}_{dmn} - \mu_d)(\mathbf{V}_{cmn} - \mu_c)}{\sqrt{(\sum_m \sum_n (\mathbf{V}_{dmn} - \mu_d)^2)\,(\sum_m \sum_n (\mathbf{V}_{cmn} - \mu_c)^2)}} \tag{8.24}$$

where, $\mathbf{V}_d = \mathbf{W}_2\mathbf{H}_2$ is the estimated discriminant TFM, $\mathbf{V}_c$ is the TFM of the chirp signal ($y_c(t)$), and μ_d and μ_c are the mean of $\mathbf{V}_d$ and $\mathbf{V}_c$ matrices, respectively. Fig. 8.11 shows that when the ChSR decreases to -40dB, the DTFM method is still able to accurately localize more than 60% of the difference in the class B signal.

As shown in this example, the proposed DTFM method successfully identified the chirp as the discriminant pattern between two signals. In the next example, we explore if the bases identified by DTFM efficiently locate the discriminant pattern of interest in a signal.

8.5.2 Localization of Region of Discrimination (ROD)

In this example, we explore if the identified discriminant bases can be used to locate the ROD in a signal. The definition of ROD may vary depending on the application; for example, in image processing, ROD is referred to the boundaries of an object of interest, or in brain disorder screening, the location of an epilepsy in an Electroencephalogram (EEG) signal is the region of interest. In pattern recognition, ROD generally refers to the regions that discriminate a signal from the signals from the other classes. Localization of the ROD could have substantial benefits in many applications.

We create three signals: Signals A and B that are selected from two different classes and signal C, which is constructed by combining signals A and B. There are two stages in this example: In the first stage, we use DTFM to identify the discriminant bases between signals A and B. In the second stage, we use these bases to locate the discriminant pattern we created in signal C.

Figs. 8.12 (a) and (b) show A and B signals. Signal A is a 3s duration of a rock music with sampling frequency of 44.1 kHz, and signal B is a 3s segment of a classic music. First, we apply the NMFDB method to the TF matrices of signal A and signal B to derive the discriminant bases in rock and classic signals. Next, we combine rock and classic frames to construct signal C. Figs. 8.12 (c) and (d) show signal C and its combination pattern, respectively. The white areas represent the rock music and the black areas show the classic music. Finally, we apply NMF to decompose the TFM of signal C into its spectral and temporal bases, and use the discriminant bases obtained from the previous stage to look for the bases that represent the rock or the classic pattern. A minimum squared error-based comparison is used to quantify the distance between bases. Once the rock and classic bases are identified, we use them to localize the pattern in signal C. Fig. 8.12 (f) shows the recognized pattern. The locations where the patterns in Figs. 8.12 (d) and (e) overlap are the areas in which we successfully recognized the rock and the classic pattern in signal C. As can be seen in this figure, the identified discriminant bases are strong enough to correctly localize a majority of the patterns in signal C.

This example showed the potential of the new DTFM method for identification and localization of the discriminant structures in signals. The next example investigates the implication of the

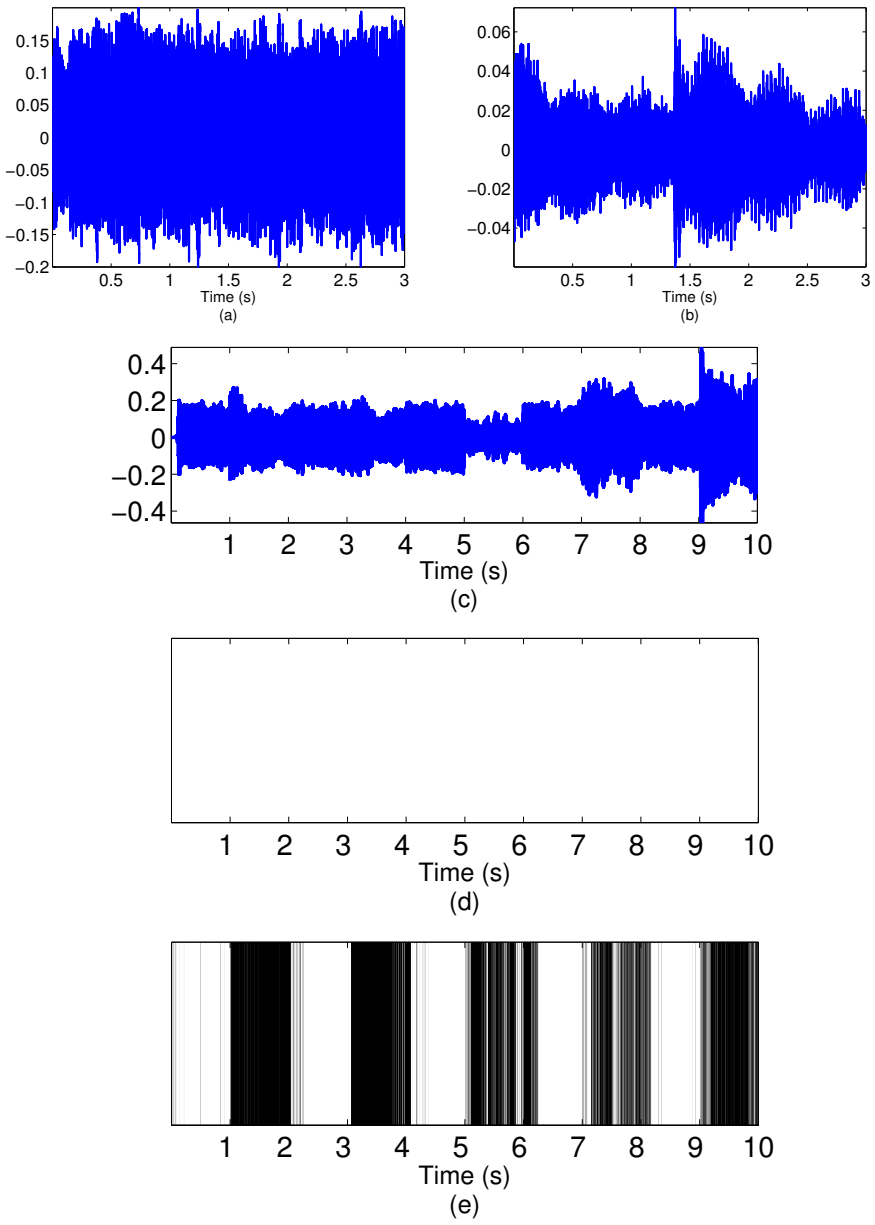


Figure 8.12: Localization of the discriminant pattern in a signal using the new DTFM method. (a) 3 s signal selected from a Rock music. (b) 3 s signal segmented from a classic music. (c) A 10 s signal generated by combining 1 s duration of rock and classic segments. (d) The rock and classic pattern in the 10 s signal; the white and the black areas show the rock and classic music, respectively. (e) The recognized pattern obtained using the DTFM method.

DTFM approach for signal classification.

8.5.3 Classification

The DTFM approach was applied to classify pathological voice disorder, which is a non-stationary biomedical signal database. Dysphonia or pathological voice refers to speech problems resulting from damage to or malformation of the speech organs. Pathological voice disorder is more common in people who use their voice professionally, for example, teachers, lawyers, salespeople, actors, and singers, and it dramatically effects these professional groups's lives both financially and psychosocially. The purpose of this work is to help patients with pathological problems for monitoring their progress over the course of voice therapy. In Chapter 7, we applied NMF to TFM decomposition of speech signals to automatically identify and measure the speech pathology problem. We proposed a new unsupervised clustering scheme that separates the abnormality discriminant features from the common features in the feature space. In the present chapter, we use our developed DTFM technique to combine the feature extraction and clustering stages, and automatically obtain the discriminant features in one stage.

The DTFM method was applied to the Massachusetts Eye and Ear Infirmary (MEEI) voice disorders database, distributed by Kay Elemetrics Corporation [84]. The database consists of 51 normal and 161 pathological speakers whose disorders spanned a variety of organic, neurological, traumatic, and psychogenic factors. The speech signal is sampled at 25 kHz and quantized at a resolution of 16 bits/sample. In this study, 25 abnormal and 25 normal signals were used to train the classifier, and then the trained classifier is used to classify the whole database. This procedure is explained as follows:

Training

This stage builds upon DTFM decomposition and feature extraction. The details of these steps are explained below:

DTFM Decomposition: DTFM decomposition identifies the discriminant bases of normal and pathological speech signals. First, the DTFM method is applied to each 0.5 s segment of a patho-

logical and a normal speech signal. In this application, spectrogram, FFT size of 1024 points and Kaiser window with parameter of five, length of 256 samples and 220 samples overlap, is used to construct the TFM of each segment. We apply NMFDB to each pathological and normal segments in the train database, and identify three discriminant bases for each pathological segments and three discriminant bases for each normal segments. Fig. 8.13 shows the above procedure for one normal and one pathological segment. The TF matrices of these two segments are shown in Figs. 8.13 (b) and (d), respectively, and the decomposed components are shown in Figs. 8.13 (e), (f) and (g). As can be observed in these figures, NMFDB successfully identified the discriminant TF structures in the normal and pathological segments. The pathological discriminant bases (Fig. 8.13 (f)) present weak formants, while the normal discriminant bases (Figure 8.13 (e)) has more periodicity in low frequencies, and introduces stronger formants. Also, the pathological discriminant bases contain the noisy structure in the pathological speech segments. The common bases (Fig. 8.13 (g)) include the joint TF structures in the pathological and the normal signal.

Feature Extraction: Eight features are extracted from each base and coefficient pair. These features are introduced in Section 5.3, and are as follows:

- Sparsity:

  Sparsity of each coefficient vector is quantified as below:

$$\mathrm{S}_{h_i} = \frac{\sqrt{N} - \left(\sum_{n=1}^{N} h_i(n)\right) / \sqrt{\sum_{n=1}^{N} h_i^2}}{\sqrt{N} - 1}, \tag{8.25}$$

  The above function is unity if and only if h_i contains a single non-zero component, and is zero if and only if all the components are equal.

- Moments:

  We extract the first two moments of each base and coefficient vectors:

$$\mathrm{MO}_W^{(o)} w_i = \sum_{m=1}^{M} f^o w_i(m), \tag{8.26}$$

$$\mathrm{MO}_H^{(o)} h_i = \sum_{n=1}^{N} t^o h_i(n), \tag{8.27}$$

$$o = 1, 2 \tag{8.28}$$

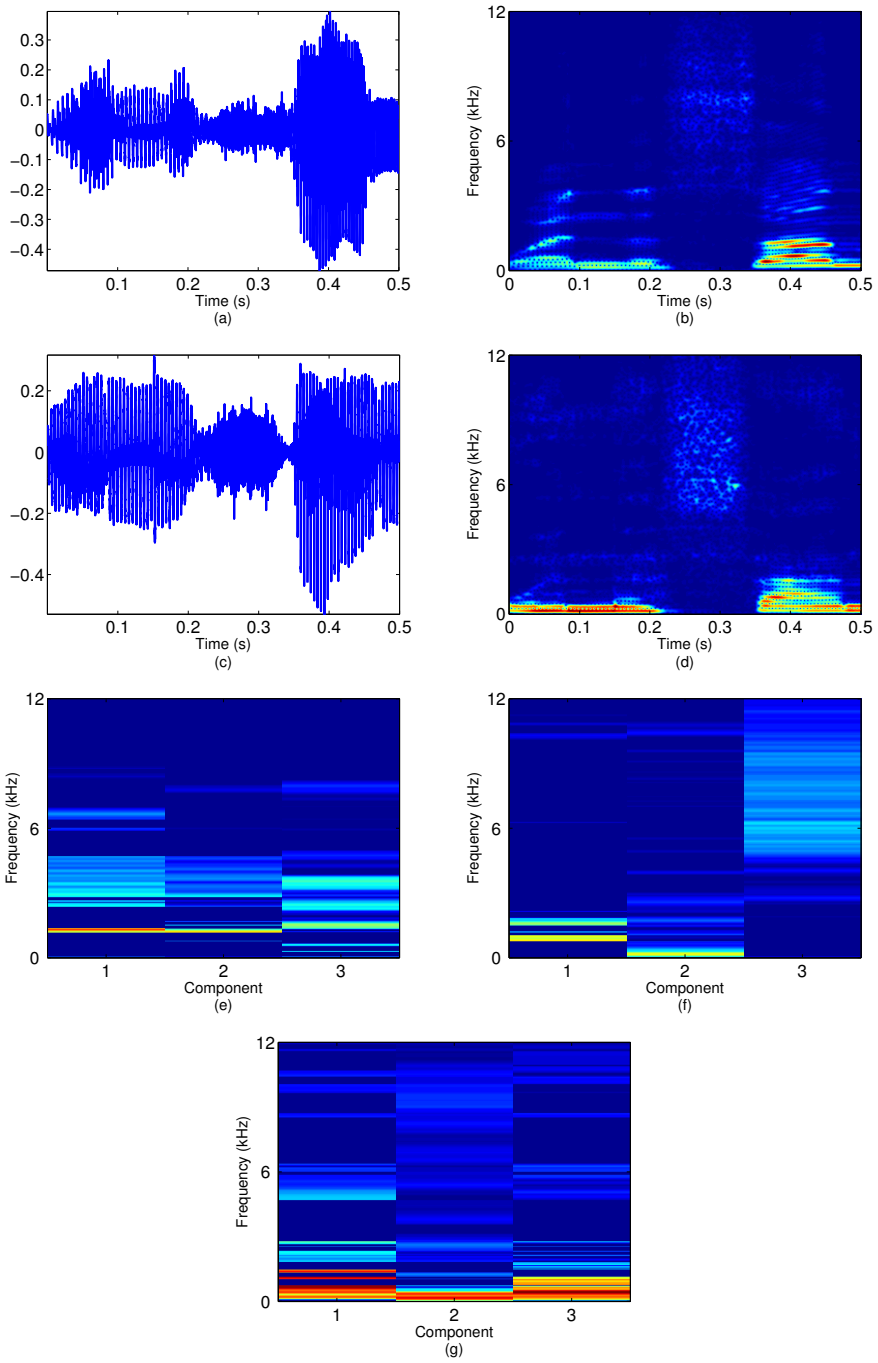


Figure 8.13: The DTFM method detects the discriminant and common bases of a pathological and a normal subject. (a) A 0.5 s segment of a normal subject. (b) TFM of the segment shown in (a). (c) A 0.5 s segment of a pathological voice disorder subject. (d) The TFM of the pathological subject shown in (c). (e) Normal discriminant bases. (f) Pathological discriminant bases. (g) Common bases.

In the above equation, $\text{MO}_W^{(1)}$ and $\text{MO}_W^{(2)}$ are the two spectral moments, and M is the frequency resolution. $\text{MO}_H^{(1)}$ and $\text{MO}_H^{(2)}$ are the temporal moments, and N is the number of samples in time.

Classifier Learning: Our focus in this study is to investigate if the new DTFM is an effective method for classification of pathological signals. Therefore, we avoid complex classifiers such as Neural Networks and Kernels, and apply a linear discriminant analysis (LDA) as a simple linear classifier for the classification stage. The NMFDB method identified three categories of base and coefficient vectors: discriminant to the normal speech signals, discriminant to the pathological voice disorder signals and common between the normal and the pathological speech signals. The extracted features from each of the three categories are fed into an LDA classifier, and a 3-class LDA classifier is trained.

Validation

In the validation stage, we use the trained classifier to classify the database. NMF with decomposition order of six ($r = 6$) is applied to the TFM of each 0.5 s segments of the signals in the database. The eight features that were explained in the training stage are extracted from each decomposed vectors. Depending on the extracted features, the trained LDA classifier decides whether each segment belongs to the normal, pathological or common class. We count the number of normal and pathological components present in each signal, and depending on which of these two categories outnumbers, we decide whether the signal belongs to a normal subject or a subject with pathological voice disorder. We applied the above procedure to all the signals in the database, and an overall classification accuracy of 97% was acheived. Table 8.1 shows the details of the classification result. From the table, it can be observed that out of 51 normal signals, 48 were classified as normal, and only 3 were misclassified as pathological. Also, the table shows that out of 161 pathological signals, 157 were classified as pathological and only 3 were misclassified as normal. The total classification accuracy is 97% which is comparable to 98.6% accuracy that we achieved in our previous work [18] where we used an unsupervised classifier to separate the discrmiminant features of the normal and pathological speech signals from the common features.

Table 8.1: Classification result.

Classes	Normal	Pathological	Total
Normal	48	3	51
Pathological	3	158	161
Normal	94.1%	5.9%	100%
Pathological	1.9%	98.1%	100%

Table 8.2 lists the accuracy rates of the several available techniques with the same pathological dataset used in this dissertation [84]. The first three rows represent the classification accuracies of

Table 8.2: Summary of several research works on voice pathology detection.

Techniques	Features	Classifier	Accuracy
DTFM Features*	DTFM Quantification	LDA	96%
TFM Features**	TFM Quantification	LDA	91%
TFM Clustered Features***	TFM Quantification	Discriminant Clustering	98.6%
Godino-Llorente [149]	MFCC	Neural Network	96%
Hadjitodorov [155]	Perturbation, Noise	Vector Quantization, LDA	92.7%
Maguire [156]	Perturbation, Noise, MFCC	LDA	87.16%
Marinaki [157]	Linear prediction coefficients	LDA	85%
Parsa [152]	Fundamental Frequencies	Linear Prediction (LP)	96.5%
Umapathy [153]	Adaptive TF Transformation (ATFT)	LDA	93.4%
Wester [158]	Harmonics-to-noise Ratio	Hidden Markov models	65%

* Chapter 8, ** Chapter 6, *** Chapter 7.

NMF-based TFM quantification, discriminant TF feature clustering, and discriminant TFM quantification approaches, respectively. The remaining rows in this table list the performance of some of the available works in literature. As evident in this table, although we used a very simple classifier in our proposed pattern classification approaches, their performances were significantly higher than the accuracy rates reported in the state-of-the-art techniques. TFM clustered features provided the highest classification rate (%98.6), which is over 7% improvment compared to the TFM features (ie. accuracy rate of 91%). This observation demonstrates our contribution in selection of the discriminant clusters to enhance the quality of the selected features. Additionally, DTFM features offered 5% increase in the accuracy rate of TFM features. It worths to mention that although TFM clustered features resulted in higher classification rate compared to DTFM technique,

selection of the parameters in the former approach is more case-sensitive, and requires more adaptation to the application's nature. On the other hand, DTFM quantification not only demonstrates a high representation and discrimination of the extracted features, but also requires less parameter tuning. Therefore DTFM quantification has the potential to be a powerful and universal tool to different pattern classifiction applications, including automatic detection of voice disorders and to voice quality assessment.

8.6 Chapter Summary

Fig. 8.14 displays the contribution flowchart, and highlights the achievement of this chapter. As our goal to extract representative and discriminative TF features to enhance the performance of pattern recognition systems, previous chapter proposed a discriminant feature clustering framework. This technique was performed on extracted TF features as a post-processing stage to identify the discriminant features. In this chapter, we proposed another methodology to fulfill the discrimination objective. The proposed framework was a novel discriminant TF quantification method that adaptively identified the long-term and discriminant TF structures between two signals to improve the detection accuracy of discriminant structures. Our studies showed that in many real-world applications, the nature of signals to be classified are very similar. However, current approaches assume that the structures of signals from these two groups are completely different, and obtain underlying structure of pathology signals without considering normal signals. Our experience evidenced that pathology characteristics that were found using these approaches included both the pathology and the common structures, and therefore, did not effectively and accurately represent the pathology activities. In the previous chapter, we improved the machine learning stage based on the proposed discriminant feature clustering technique to obtain the key discriminant features. The present chapter looked over the general feature extraction methodology, and presented a new framework that automatically identified the differences between signals as part of the TF quantification stage rather as a post-processing tool. It then used the identified structure to accurately detect the pathological structure in the signal. The proposed methodology is an emerging technique in biomedical and multimedia engineering, and has the potential to be a powerful and useful

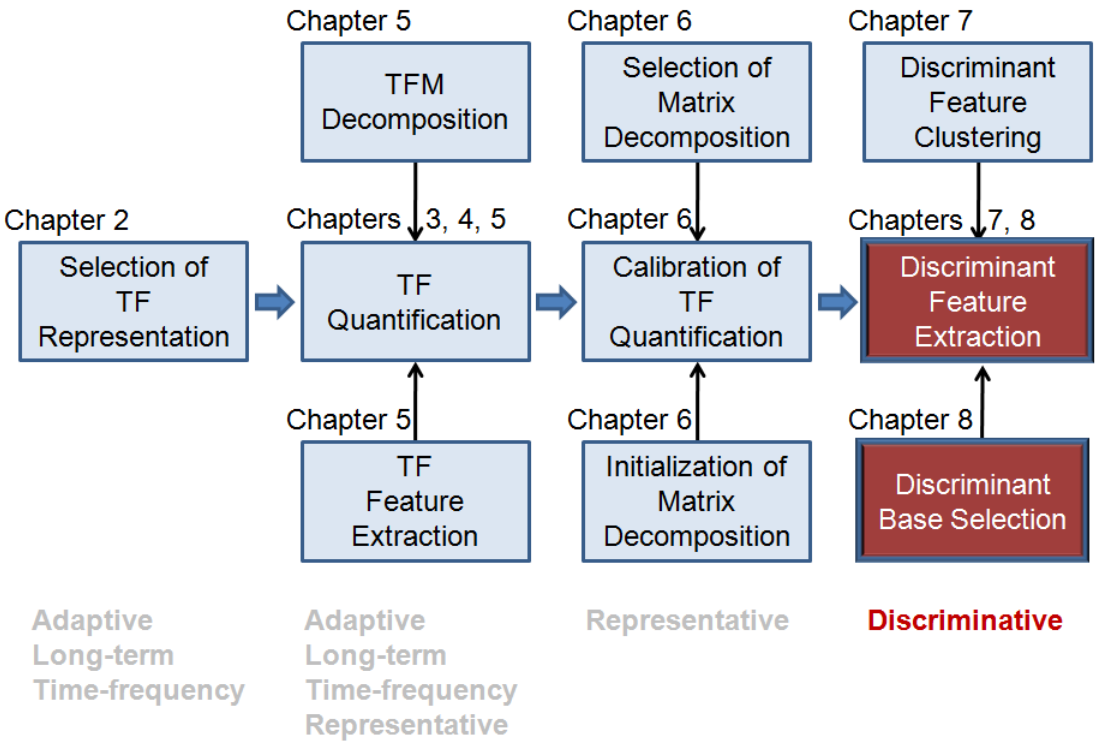


Figure 8.14: Flowchart of the proposed contributions.

tool to accurately diagnose and monitor activities of interest in advanced technologies.

In order to achieve such a discriminant TF quantification mentioned above, we modified the TFM decomposition method in a way that it flexibly identifies the discriminant bases of each class. At the first stage, we proposed a new NMF discriminant bases (NMFDB) technique which was applied to the TF matrices of two signals from two different classes, and derived the discriminant TF bases in each signal. The visualization of the new NMFDB decomposition method for synthetic and speech signals produced desirable results demonstrating the effectiveness of the NMFDB method. A quantification approach is suitable for pattern recognition purposes and is robust to amplitude scaling and temporal shift. We demonstrated that NMFDB satisfies these two desired properties.

At the second stage, discriminant features were calculated from each discriminant TF base obtained above. The features extracted from the DTFM method provided not only a better representation of each class, but also, a greater discrimination between the classes. Experiments performed with real-world and synthetic signals demonstrated the potential of the proposed novel DTFM approach as a strong TF quantification tool in the areas of i) detection of the discrimination pattern; ii) localization of the region of discrimination (ROD); and iii) feature extraction and classification.

Chapter 9

CONCLUSION

AN adaptive signal processing framework for efficiently analyzing and extracting features from non-stationary signals was presented in this dissertation. The proposed method was evaluated using synthetic and real world signals in different stages, and desirable results were achieved. Fig. 9.1 displays the contribution flowchart as evolved throughout this dissertation.

Chapter 1 presented a detailed introduction on non-stationary signal analysis and pattern recognition, and the constraints of feature extraction techniques. In Chapter 2, we presented various TF approaches and explained the desirable properties of a suitable representation considering TF quantification. We justified the choice of the Adaptive TF approach based on its desirable properties and achievable TF resolutions. The remaining chapters focused on TF quantification to extract representative and discriminative TF features to enhance the performance of pattern recognition systems. TF feature extraction of signals with known structures were explored in Chapter 3. Adaptive TF feature extraction to successfully quantify and detect patterns of interest, and the adaptabilities to track the signals' non-stationarities were discussed in the same chapter. In Chapter 4, we presented DPPT as an efficient tool to quantify signals with embedded patterns. The extracted TF features successfully quantified the time-varying frequency of embedded patterns. The DPPT-based technique accurately detected the embedded pattern even in the presence of severe noise in the signal.

Chapters 3 and 4 investigated signal processing techniques, analyzing known TF feature structures. There are real-world applications in which the structures of interest are not known as a prior

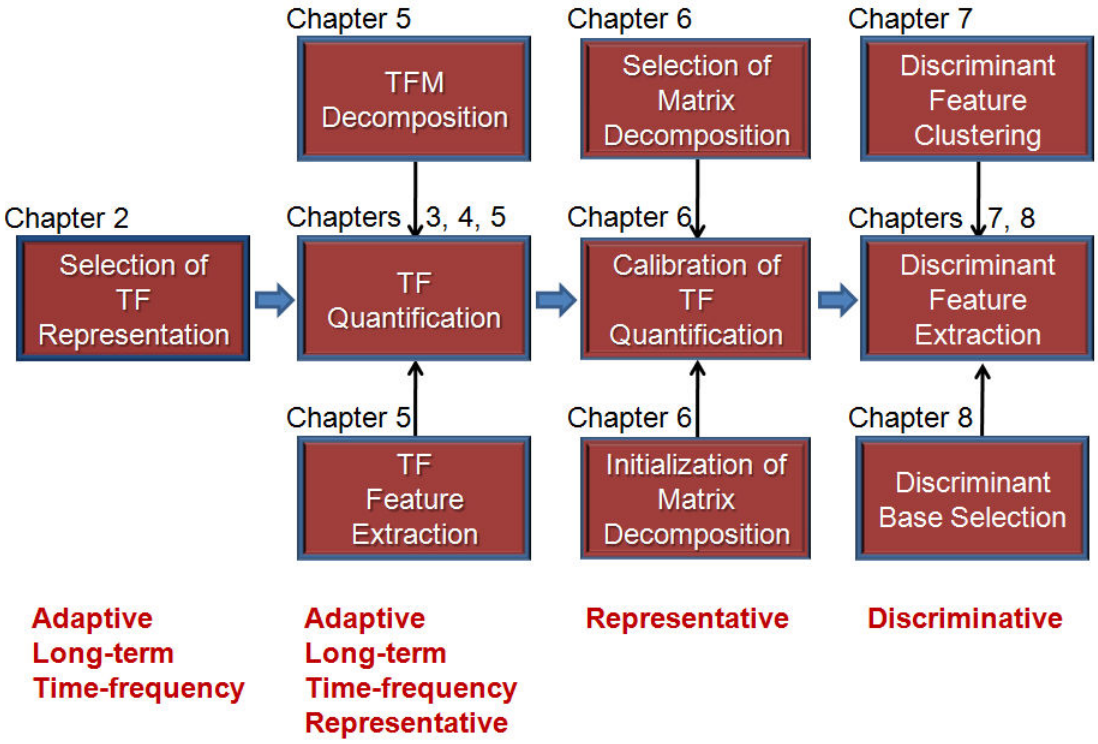


Figure 9.1: Flowchart of the proposed contributions.

knowledge. Therefore, in Chapters 5 to 8 of this dissertation, we focused on quantification of TF features with any unknown complex structure. A novel TFM quantification was introduced in Chapter 5, and its suitability for non-stationary signal analysis was discussed with synthetic signal examples. Chapter 6 discussed the various existing matrix decomposition (MD) techniques and justification in arriving at a suitable MD decomposition approach for TF analysis. Few modifications in the MD optimization process were proposed to enhance the performance of the TFM quantification. Several synthetic and real-world applications presented to verify the effectiveness of the proposed TFM quantification framework. Next, a novel discriminant feature clustering framework was proposed in Chapter 7. Finally, we introduced our novel work in discriminant base selection in Chapter 8, and presented the multifold benefits of the proposed work with synthetic and real signal examples.

9.1 Outcome of the proposed work

Table 9.1, summarizes the various solutions provided by the proposed adaptive signal processing framework in efficiently analyzing non-stationary signals and extracting long-term and discriminative features from them. The proposed framework with discrete TFM (DTFM) quantification as its significant highlight is expected to become a versatile non-stationary signal analysis tool, which has the benefits of TFM and discriminant analyses. The outcome of the proposed work could be grouped into two core contributions:

9.1.1 Core Theoretical Contributions

The following summarizes our core theoretical contributions in the overall area of TF feature analysis.

Time-frequency Matrix Quantification

Our main contribution in signal processing stage focuses attention on developing a long-term and discriminative TF analysis. To fulfill this objective, in the first point, a time-frequency matrix (TFM) decomposition was proposed to increase the effectiveness of segmentation in real-world

Table 9.1: Summary of the proposed solutions and the requirement for efficient non-stationary signal analysis

Requirements for efficient non-stationary signal analysis and feature extraction	**Solution provided/ suggested by the proposed work**	**Chapter Reference**
TF analysis	Adaptive and high resolution TFD	Chapter 2
Known TF feature detection	Adaptive TF feature extraction	Chapter 3
Embedded TF feature detection	DPPT-based feature extraction	Chapter 4
Unknown TF feature detection	Adaptive TF quantification	Chapters 5, 6, 7 and 8
Long-term signal processing	TFM decomposition	Chapter 5
Instantaneous and localized features	TFM feature extraction	Chapter 5
Representative and meaningful features	Selection of NMF	Chapter 6
Improvement of MD optimization	Integration of TFD and TFM quantification: MP-based seeding	Chapter 6
Discriminative features	Unsupervised feature clustering The Soft and Fuzzy Supervised Labellings	Chapter 7
Discriminant bases	The DTFM decomposition The NMFDB method	Chapter 8
ROD Localization	The DTFM decomposition	Chapter 8

signals. The key insight behind this approach was that in many applications, activity of interest contains non-stationary behaviours, such as transients and discontinuities that cannot be captured in traditional short-term approaches. The proposed TFM framework resolved the shortcomings of short-term analysis by adaptively splitting any signal into its stationary parts. The new TFM analysis tool improved the accuracy and effectiveness of signal quantification and detection, and suggested a novel research direction in advanced health and multimedia technologies. In the second point, meaningful and unique features were extracted from the decomposed TF components. Conventional feature extraction techniques diminished the localization properties of the instantaneous TF features by calculating some statistical properties, such as, mean and variance. In contrast to the classic approaches, we extracted meaningful features that successfully represented the instantaneous spectral and temporal structures of the given data. The calculated features were robust to noise and outliers and were successfully used for classification and localization of the discriminant patterns of signals.

Matrix decomposition (MD) selection

In literature, performance comparison of well-known MD methods has been investigated for different applications. However, depending on the application and the applied database, contradictory results have been reported. We performed a fair comparison of the MD techniques for the quantification of the TF plane, and selected the most suitable MD method for the proposed TFM decomposition technique.

TF matrix decomposition (TFMD) initialization

Our motivation in the proposed MD seeding method was to use the knowledge in the TF structure of a signal to find suitable initialization values for the decomposed matrices. The selected MD algorithm starts with a random initialization for the decomposed matrices, and modifies them iteratively until a cost function is minimized. However, due to the non-convexity of the cost function, depending on the initial matrices at each optimization, a different local minima of the cost function may be achieved. The proposed method integrated the TF analysis into the MD technique, and offered an improved seeding method for the MD optimization.

Discriminant TFM Decomposition

Based on the above TFM technique, a unique and novel discriminant TF analysis method was proposed to perform automated and discriminative feature selection of any non-stationary signal. Classic feature extraction methods calculated features from each class without considering the structure in other classes. Therefore, the feature space included overlapping features that limited the method to effectively and accurately represent the discriminative structure in each class. In order to address this problem, we proposed the discriminant TFM (DTFM) framework, which is a combination of TFM decomposition and unsupervised clustering techniques. DTFM automatically identifies the differences between different classes as the distinguishing structure, and uses the identified structure to accurately classify and locate the discriminant structure in the signal. The proposed methodology is an emerging technique in non-stationary signal analysis, and has the potential to be a powerful and useful tool to accurately diagnose and monitor activities of interest in advanced health technologies.

NMF Discriminant base (NMFDB)

In order to achieve such a discriminant TF quantification framework, we modified the TFM decomposition method in a way that it flexibly identified the discriminant bases of each class. We developed a new NMF discriminant bases (NMFDB) technique, which was applied to the TF matrices of two signals from two different classes, and derived the discriminative TF bases in each signal. The visualization of the new NMFDB decomposition method for synthetic and real signals produced desirable results, demonstrating the effectiveness of the NMFDB method. The NMFDB technique was invariant to signal translations such as amplitude scaling and temporal shift.

Discriminant Feature Clustering

A new discriminant clustering approach was offered to improve the classification accuracy in automated decision making systems. A feature clustering technique is developed through a new machine learning approach that automatically identifies the clusters of key features that represent the discriminative patterns. The discriminative clusters were then used to compute the presence of

the discriminative patterns in any given signal and classified them accordingly.

9.1.2 Core Practical Contributions

Several real-world applications are employed to evaluate the proposed work. These application are as follows:

Biomedical Signal Processing

Detect of The Risk of Sudden Cardiac Death: Each year between 0.5 to 1 million North Americans and Europeans die from sudden cardiac death (SCD) caused by ventricular arrhythmias (VA). However, identifying those patients at risk of SCD remains a formidable challenge as many people are asymptomatic until the VA event occurs, and the majority do not survive the first episode. The standard method for assessing whether a patient is at risk for SCD has been an Electrophysiology (EP) study from inside the heart. However, the EP study is invasive, expensive, and entails some risk to the patient. Therefore, there is a strong need to develop a technology that is quick, non-invasive, relatively inexpensive, yet accurate in identifying those who are at high risk of VA, and benefit from the expensive therapy. In this dissertation, we applied the proposed Adaptive feature extraction technique to provide a reliable and accurate SCD prediction to replace the invasive testing to identify patients at high risk of VA. The Adaptive signal analysis technique was superior to classic techniques in terms of tracking the non-stationarity and preserving robustness in the presence of noise. Therefore, this technique has a high potential of technology transfer that may lead to the development of novel implications in cardiac monitoring that can benefit global health care.

Pathological Speech Recognition In the past 20 years, a significant attention has been paid to the science of voice pathology diagnostic and monitoring. The purpose of this work is to help patients with pathological problems for monitoring their progress over the course of voice therapy. Currently, patients are required to routinely visit a specialist to follow up their progress. Moreover, the traditional ways to diagnose voice pathology are subjective, and depending on the experience of the specialist, different evaluations can be resulted. Developing an automated technique saves time for both the patients and the specialist, and can improve the accuracy of the assessments.

In this dissertation, we used the proposed TFM quantification to classify MEEI voice disorders database, including 161 pathological and 51 normal speakers, and achieved a significant accuracy rate of 98.6%.

Multimedia Signal Processing

Multimedia Security Detection: There are about 0.5 trillion copies of sound recordings in existence and 20 billion sound recordings are added every year. This underscores the importance of securing content. One of the approaches to multimedia security is watermarking or fingerprinting, which is the process of embedding additional data into the host signal for identifying copyright ownership. In this dissertation, we used the Adaptive TF feature extraction approach to improve the robustness of the security process under the intentional or unintentional signal manipulations.

Environmental Audio Scene Analysis: Audio signals are important sources of information for understanding the content of multimedia. Therefore, developing audio classification techniques that better characterize audio signals plays an essential role in many multimedia implications, such as, multimedia indexing and retrieval, and auditory scene analysis. In this dissertation, we demonstrated that the proposed adaptive TF signal analysis framework significantly enhances the pattern classification of audio signals.

9.2 Limitations and Future Work

The following would be the directions for future work in applying and enhancing the proposed work with more intelligence and accuracy.

- DTFM quantification exhibit desirable properties to detect discriminating signal patterns in a two-group classification problem. This idea could be extended to design a novel multi-dimensional adaptive system that could identify the patterns of discrimination in scenarios with more than two groups. The optimization of NMFDB could be modified to decompose the TF matrices from different sources in a way that the discriminant bases are separated.

- Although the proposed TF analysis is applied on single channel signals, the approach can

be extended to multi-channel biomedical signals. The long-term and discriminating nature of the DTFM quantification frameworks could be converted into a 3D-DTFM to quantify the deviation of the biosignals and attribute the deviation to possible causes of a medical situation. Quantifying the discriminant patterns could be used to localize the biomedical signals of interest from other sources or artifacts.

- The developed DTFM quantification framework has the potential to a powerful and universal signal analysis tool to different pattern classification applications which involve complex and non-stationary signal structures. In order to arrive at such a universal template for extraction of the discriminant features, extensive training and calibration of the DTFM quantification is needed. The parameters in the NMFDB technique are required to be optimized in order to obtain the most efficient discriminant quantification. This research would result in the development of an adaptable and universal non-stationary signal analysis for a variety of signals.

- The NMF technique decomposed a TFM into r spectral and temporal components where the values of r was experimentally selected. The more accurate we decompose the TFM into its components, the more accurate we could extract the discriminatory features. In Chapter 6, we used the decompositions resulted from the Adaptive TF representation to improve the NMF seedings. This improvement can be extended to search for the optimum decomposition number using the structure of signals in Adaptive TF representation.

- The advantages of the proposed technique outweigh the computational expenses. In the near future, the fast growing technological developments will significantly reduce the processing time so that the technique will eventually become a real-time processing tool. Additionally, the technique can be implemented in a dedicated device to be more easily marketable especially in the health care and other related applications.

List of Publications

Published in Refereed Journals and Proceedings

1. B. Ghoraani and S. Krishnan, "A Joint Time-Frequency and Matrix Decomposition Feature Extraction Methodology for Pathological Voice Classification", The EURASIP Journal on Advances in Signal Processing, vol. 2009, Article ID 928974, 11 pages, 2009, doi:10.1155/2009/928974

2. K. Umapathy, B. Ghoraani and S. Krishnan, Audio Signal Processing using Time-frequency Approaches: Coding, Classification, Fingerprinting, and Watermarking, to appear in the EURASIP Journal on Advances in Signal Processing, ASP/451695, Feb. 2010

3. B. Ghoraani, and S. Krishnan, Discriminative Base Decomposition for Time-frequency Matrix Decomposition", in the proceedings of the 35th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010), March 2010,

4. Nasim Shams, B. Ghoraani and S. Krishnan, " Audio Feature Clustering for Hearing Aid Systems", in the proceedings of the IEEE Toronto International Conference - Science and Technology for Humanity (TIC-STH 2009), Sept. 26-27, 2009, Page(s): 976-680, Toronto, Canada

5. B. Ghoraani, S. Krishnan, R. J. Selvaraj and V. S. Chauhan, " Adaptive Time-frequency Matrix Features for T wave Alternans Analysis ", Invited paper, in the proceedings of the 31st IEEE Eng in Medicine and Biology Society Conf (EMBC 2009), Page(s): 39 42 September 2-6, 2009, Minneapolis, Minnesota, USA

6. B. Ghoraani, S. Krishnan, R. J. Selvaraj and V. S. Chauhan, "Adaptive Time-Frequency Signal Analysis and its Case Study in Biomedical ECG Waveform Analysis", in the proceedings of the 16th International Conference on Digital Signal Processing (DSP 2009), Page(s): 1 5, July 5-7 2009, Santorini, Greece

7. B. Ghoraani and S. Krishnan , ”Quantification and localization of features in time-frequency plane”, the proceedings of the IEEE Canadian Conf on Electrical and Computer Engineering (CCECE 2008), 4-7 May 2008, Niagara Falls, Ca, Page(s):1207 1210

8. S. Krishnan, B. Ghoraani, and S. Erkucuk, ”Time-frequency Analysis of Digital Audio Watermarking ”, Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, Information Science, Reference ISBN: 978-1-59904-513-9, Hershey, PA, 17033-1240, USA, 2007

9. Ghoraani, and S. Krishnan, ”Chirp-based image watermarking as error-control coding”, in the proceedings of the IEEE Intern Conf on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2006), Dec. 2006, Pasadena, Page(s): 647 650

10. L. Le, S. Krishnan, and B. Ghoraani, ”Discrete Polynomial Transform for Digital Image Watermarking Application”, in the proceedings of the IEEE Intern Conf on Multimedia and Expo (ICME 2006), July 2006, Toronto, Ca, Page(s). 1569 1572

Submitted or to be Submitted to Refereed Journals

11. B. Ghoraani and S. Krishnan, Time-Frequency Matrix Feature Extraction and Classification of Environmental Audio Signals, to the IEEE Trans. Audio, Speech and Language Processing.

12. B. Ghoraani and S. Krishnan, T Wave Alternans Analysis Using Adaptive Time-Frequency Signal Analysis and Non-negative Matrix Factorization”, to Medical Engineering and Physics.

13. B. Ghoraani and S. Krishnan, Discrete Bases Selection in Time-frequency Matrix Analysis, to the IEEE Transactions on Signal Processing.

14. B. Ghoraani and S. Krishnan, Discriminant Feature Clustering, An Integration of Unsupervised and Supervised Machine Learning, to the IEEE Trans. Pattern Analysis and Machine Intelligence.

15. B. Ghoraani and S. Krishnan, A Comprehensive Analysis of Time-Frequency Matrix Quantification, to Signal, Image and Video Processing (Springer).

Bibliography

[1] "The national space science and technology center," *Data: http://vortex.nsstc.uah.edu/public/msu/t2lt/tltglhmam-5.2*. [Online]. Available: http://vortex.nsstc.uah.edu/

[2] M. Akay, "Time frequency and wavelets in biomedical signal processing," *IEEE Press*, vol. ISBN 0780311477, 9780780311473, 1997.

[3] E. Duzel, R. Habib, B. Schott, A. Schoenfeld, N. Lobaugh, A. R. McIntosh, M. Scholz, and H. J. Heinze, "A multivariate, spatiotemporal analysis of electromagnetic time-frequency data of recognition memory," *Neuroimage*, vol. 18, no. 1, pp. 185–97, January 2003.

[4] W. Williams, H. Zaveri, and C. Sackellares, "Timefrequency analysis of electrophysiology signals in epilepsy," *IEEE Eng Med Biol*, vol. 14, no. 2, p. 13343, 1995.

[5] M. Stridh, L. Sornmo, C. Meurling, and S. Olsson, "Characterization of atrial fibrillation using the surface ecg: time-dependent spectral properties," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 19–27, January 2001.

[6] E. M. Bernat, W. J. Williams, and W. J. Gehring, "Decomposing erp timefrequency energy using pca," *Clinical Neurophysiology*, vol. 116, no. 6, pp. 1314–1334, June 2005.

[7] A. Delorme, S. Makeig, M. Fabre-Thorpe, and T. Sejnowski., "From single-trial eeg to brain area dynamics," *Neurocomputing*, vol. 44-46, p. 10571064, 2002.

[8] M. Mrup, L. Hansen, J. Parnas, and S. M. Arnfred, "Decomposing the time-frequency representation of eeg using nonnegative matrix and multi-way factorization,"

Technical report, Informatics and Mathematical Modeling, Technical University of Denmark, 2006. [Online]. Available: http://www2.imm.dtu.dk/pubdb/views/edoc download.php/4144/pdf/imm4144.pdf

[9] T. M. Rutkowski, R. Zdunek, and A. Cichocki, "Multichannel eeg brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support," *In the Proceedings of International Congress Series*, vol. 1301, p. 266269, 2007.

[10] F. Miwakeichi, E. Martinez-Montes, P. Valds-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing eeg data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[11] B. Tacer and P. Loughlin, "Time-frequency based classification," *In Proceedings of the International Society for Optical Engineering (SPIE)*, vol. 2846, pp. 186–192, August 1996.

[12] ——, "Nonstationary signal classification using the joint moments of time-frequency distributions," *Pattern recognition*, vol. 39, pp. 419–424, 1998.

[13] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," *Computers in Biology and Medicine*, vol. 34, no. 6, pp. 523 – 537, 2004.

[14] P. Smaragdis and J. Brown, "Non-negative matrix factorization for polyphonic music transcription," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177– 180, October 2003.

[15] K. Englehart, B. Hudgins, P. Parker, and M. Stevenson, "Classification of the myoelectric signal using time-frequency based representations," *Medical Engineering and Physics - Elsevier*, vol. 21, no. 6, pp. 431–438, 1999.

[16] H. Kim, J. Burred, and T. Sikora, "How efficient is mpeg-7 for general sound recognition?" *In the proceedings of 25th International AES Conference, London, U.K.*, 2004.

[17] A. Holzapfel and Y. Stylianou, "Musical genre classification using nonnegative matrix factorization-based features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 424–434, February 2008.

[18] B. Ghoraani and S. Krishnan, "A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. ID 928974, pp. 11 pages, doi:10.1155/2009/928 974, 2009.

[19] ——, "Quantification and localization of features in time-frequency plane," *Proceedings of IEEE CCECE 2008*, pp. 1207–1210, May 2008.

[20] D. Groutage and D. Bennink, "Feature sets for nonstationary signals derived from moments of the singular value decomposition of cohen-posch (positive time-frequency) distributions," *EEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1498–1503, May 2000.

[21] ——, "A new matrix decomposition based on optimum transformation of the singular value decomposition basis sets yields principal features of time-frequency distributions," *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing*, vol. 48, pp. 598–602, Augest 2000.

[22] N. Saito and R. Coifman, "Local discriminant bases and their applications," *Jornal of Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 337–358, 1995.

[23] L. Deqiang, W. Pedrycz, and N. Pizzi, "Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 6, pp. 1132–1139, June 2005.

[24] K. Umapathy, S. Krishnan, and R. Rao, "Audio signal feature extraction and classification using local discriminant bases," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1236–1246, May 2007.

[25] S. Davis, B. D. V. Veen, S. C. Hagness, and F. Kelcz, "Breast tumor characterization based on ultrawideband microwave backscatter," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 237–246, January 2008.

[26] N. F. Ince, F. Goksua, A. H. Tewfika, and S. Aricad, "Adapting subject specific motor imagery eeg patterns in spacetimefrequency for a brain computer interface," *Biomedical Signal Processing and Control, New Trends in Voice Pathology Detection and Classification*, vol. 4, no. 3, pp. 236–246, July 2009.

[27] K. Umapathy and S. Krishnan, "Time-Width Versus Frequency Band Mapping of Energy Distributions," *IEEE Transactions on Signal Processing*, vol. 55, pp. 978–989, Mar. 2007.

[28] H.-I. Choi and W. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 6, pp. 862 – 871, June 1989.

[29] I. Daubechies, "The wavelet transform, time-frequency localization and signalanalysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[30] Z. Peng, P. W. Tse, and F. Chu, "An improved hilberthuang transform and its application in vibration signal analysis," *Journal of Sound and Vibration - Elsevier*, vol. 286, no. 1-2, pp. 187–205, August 2005.

[31] L. Cohen and T. Posch, "Positive time-frequency distribution functions," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 31–38, January 1985.

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

[33] L. Cohen, "Time-frequency distributions – a review," *Proceedings of the IEEE*, vol. 77, pp. 941–981, 1989.

[34] S. Krishnan, R. Rangayyan, G. Bell, and C. Frank, "Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 6, pp. 773– 783, June 2000.

[35] P. Loughlin, J. Pitton, and L. Atlas, "Construction of positive time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2697–2705, October 1994.

[36] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068 – 1089, May 1995.

[37] S. Mallat, "A wavelet tour of signal processing," *Academic Press*, 1998.

[38] R. J. Selvaraj and V. S. Chauhan, "Effect of noise on t-wave alternans measurement in ambulatory ecgs using modified moving average versus spectral method," *Pacing Clin Electrophysiol*, vol. 32, pp. 632 – 641, May 2009.

[39] J. M. Smith, E. A. Clancy, C. R. Valeri, J. N. Ruskin, and R. J. Cohen, "Electrical alternans and cardiac electrical instability," *Circulation*, vol. 77, no. 1, p. 110121, 1988.

[40] J. Martinez and S. Olmos, "Methodological principles of t wave alternans analysis: a unified framework," *IEEE Transactions on Biomedical Engineering*, vol. 52, pp. 599 – 613, April 2005.

[41] L. Burattini, W. Zareba, J. P. Couderc, E. L. Titlebaum, and A. J. Moss, "Computer detection of nonstationary t-wave alternans using a newcorrelative method," *Computer on Cardiology*, vol. 24, p. 657660, 1997.

[42] L. Burattini, W. Zareba, and A. J. Moss, "Correlation method for detection of transient t-wave alternans in digital holter ecg recordings," *Annals of Noninvasive Electrocardiology*, vol. 4, no. 4, p. 416 426, 1999.

[43] B. D. Nearing and R. L. Verrier, "Modified moving average analysis of t-wave alternans to predict ventricular fibrillation with high accuracy," *Journal of applied physiology*, no. 92, p. 541549, 2002.

[44] B. D. Nearing, A. H. Huang, and R. L. Verrier, "Dynamic tracking of cardiac vulnerability by complex demodulation of the t wave," *Science*, no. 252, p. 437440, 1991.

[45] P. Strumillo and J. Ruta, "Poincar mapping for detecting abnormal dynamics of cardiac repolarization," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, p. 6265, 2002.

[46] T. Srikanth, D. Lin, N. Kanaan, and H. Gu, "Estimation of low level alternans using periodicity transformsimulation and european st/t database results," *Proceedings of IEEE Engineering in Medicine and Biology Society (EMBS)*, p. 14071408, 2002.

[47] D. Bloomfield, S. Hohnloser, and R. Cohen, "Interpretation and classification of microvolt t wave alternans tests," *Journal of Cardiovasc Electrophysiol*, vol. 16, pp. 502 – 512, 2002.

[48] G. Moody, W. Muldrow, and R. Mark, "A noise stress test for arrhythmia detectors," *Computers in Cardiology*, vol. 11, no. 3, pp. 381–384, 1984.

[49] H. Bazett, "The time relations of the blood-pressure changes after excision of the adrenal glands, with some observations on blood volume changes," *J Physiol*, vol. 53, pp. 320–339, 1920.

[50] R. Rangayyan and S. Krishnan, "Feature identification in the time-frequency plane by using the Hough-Radon transform," *IEEE Trans. on Pattern Recognition*, vol. 34, pp. 1147–1158, 2001.

[51] A. Francos and M. Porat, "Analysis and synthesis of multicomponent signals using positive time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 47, pp. 493–504, February 1999.

[52] S. Peleg and B. Friedlander, "Multicomponent signal analysis using the polynomial-phase transform," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 378–387, January 1996.

[53] ——, "The discrete polynomial-phase transform," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1901–1914, August 1995.

[54] L. Sin-Joo and J. Sung-Hwan, "A survey of watermarking techniques applied to multimedia," *IEEE International Symposium on Industrial Electronics*, vol. 1, pp. 272–277, June 2001.

[55] R. Bangaleea and H. Rughooputh, "Performance improvement of spread spectrum spatial-domain watermarking scheme through diversity and attack characterisation," *IEEE Africon Conference, Africa*, vol. 1, pp. 293–298, October 2002.

[56] S. Yafei, z. Li, W. Guowei, and L. Xinggang, "A novel frequency domain watermarking algorithm with resistance to geometric distortions and copy attack," *In the Proceedings of the International Symposium on Circuits and Systems, Thailand*, vol. 2, pp. II–940–II–943, May 2003.

[57] B. Mobasseri, "Digital watermarking in joint time-frequency domain," *Proc. of International Conference on Image Processing, New York, USA*, vol. 3, pp. II–481–II–484, June 2002.

[58] D. Kundur and D. Hatzinakos, "Toward robust logo watermarking using multiresolution image fusion principles," *IEEE Transactions on Multimedia*, vol. 6, pp. 185–195, February 2004.

[59] A. A. Reddy and B. N. Chatterji, *Source Pattern Recognition Letters,'A new wavelet based logo-watermarking scheme'*. New York: Elsevier Science Inc., May 2005, vol. 26, no. ISSN:0167-8655.

[60] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Information Theory*, vol. 6, no. 12, pp. 1673–1687, December 1997.

[61] I. Cox, M. Miller, and J. Bloom, "Digital watermarking," *San Diego, CA, Academic Press*, 2002.

[62] M. Swanson, B. Zhu, and A. Tewfik, "Current state of the art, challenges and future directions for audio watermarking," *Proc. IEEE Intl. Conf. Multimedia Computing and Systems*, vol. 1, pp. 19–24, June 1999.

[63] W. Lie and L. Chang, "Robust high quality time-domain audio watermarking subject to psychoacoustic masking," *Proc. IEEE Intl.Symp. Circuits and Systems*, vol. 2, pp. 45–48, May 2001.

[64] J. Seok and J. Hong, "Audio watermarking for copyright protection of digital audio data," *Electronics Letters*, vol. 37, no. 1, pp. 60–61, January 2001.

[65] L. Gang, A. Akansu, and M. Ramkumar, "Mp3 resistant oblivious steganography," *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing*, vol. 3, pp. 1365–1368, May 2001.

[66] B. M. Macq and J. Quisquater, "Cryptology for digital tv broadcasting," *Proceedings of the IEEE*, vol. 83, pp. 944–957, June 1995.

[67] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, no. 3-4, pp. 313–336, 1996.

[68] J. M. Barton, *Method and apparatus for embedding authentication information within digital data*. United States Patent, 1997.

[69] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[70] D. Kirovski and H. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transactions on Signal Processing, special issue on data hiding*, vol. 51, pp. 1020–1034, April 2003.

[71] D. T. N. Cvejic and T. Seppanen, "Increasing robustness of an audio watermark using turbo codes," *Proc. IEEE. Intl. Conf. Multimedia and Expo*, vol. 1, pp. 217–220, July 2003.

[72] S. Erkucuk, S. Krishnan, and M. Zeytinoglu, "Robust audio watermarking using a chirp based technique," in *IEEE Intl. Conf. on Multimedia and Expo*, vol. 2, 2002, pp. 513–616.

[73] A. Ramalingam and S. Krishnan, "Robust image watermarking using a chirp detection-based technique," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 152, pp. 771–778, December 2005.

[74] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based models of human perception," *Proceedings of the IEEE*, vol. 81, p. 13851422, 1993.

[75] S. Kay and G. Boudreaux-Bartels, "On the optimality of the wigner distribution for detection," *In the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10, pp. 1017–1020, April 1985.

[76] J. Dhanoa, E. Hughes, and R. Ormondroyd, "Simultaneous detection and parameter estimation of multiple linear chirps," *ICASSP, IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. VI – 129–32, April 2003.

[77] J. Torres, P. Cabiscol, and J. Grau, "Radar chirp detection through wavelet transform," *Proceedings of the 5th Biannualg on World Automation Congress*, vol. 13, pp. 227–232, June 2002.

[78] J. Lee, H. Kim, and J. Lee, "Information extraction method without original image using turbo code," *Proc. International Conference on Image Processing, Greece*, vol. 3, pp. 880–883, October 2001.

[79] S. Pereira, S. Voloshynovskiy, M. Madueno, S. Marchand-Maillet, and T. Pun, "Second generation benchmarking and application oriented evaluation," in *Information Hiding Workshop III*, Pittsburgh, PA, USA, April 2001.

[80] X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," *Proceedings of the IEEE International Conference on Image Processing, ICIP, California, USA*, vol. 1, pp. 548 – 552, October 1997.

[81] J. R. Kim and Y. S. Moon, "A robust wavelet-based digital watermark using level-adaptive thresholding," *Proceedings of the 6th IEEE International Conference on Image Processing, ICIP, Kobe, Japan*, pp. 202 – 206, October 1999.

[82] H. Hassanpour, M. Mesbah, and B. Boashash, "Timefrequency feature extraction of newborn eeg seizure using svd-based techniques," *EURASIP Journal on Applied Signal Processing*, vol. 16, pp. 2544–2554, 2004.

[83] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, p. 14571469, 2004.

[84] E. I. M. Eye, "Voice disorders database," *Lincoln Park, NJ: Kay Elemetrics Corporation*, no. Version 1.03.

[85] M. Lee, A. Lee, D. K. Lee, and S.-Y. Lee, "Video representation with dynamic features from multi-frame frame- difference images," *In the Proceedings of the IEEE Workshop on Motion and Video Computing*, vol. 3, pp. 28–35, February 2007.

[86] I. Ciocoiu, "Occluded face recognition using parts-based representation methods," *Proceedings of the 2005 European Conference on Circuit Theory and Design*, vol. 1, pp. 315–318, 28 Augest - 2 September 2005.

[87] N. Chikhi, B. Rothenburger, and N. Aussenac-Gilles, "A comparison of dimensionality reduction techniques for web structure mining," *IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 116–119, November 2007.

[88] Y. C. Cho, S. Choi, and S. Y. Bong, "Non-negative component parts of sound for classification," *In the proceedings of IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT), Darmstadt, Germany*, pp. 633–636, 2003.

[89] J. Sylvester, "On the reduction of a biline quantic of the nth. order to form of a sum. of n products by a doubles orthogonal substitution," *Messenger of Math*, vol. 19, 4246 1889.

[90] C. Croux and G. Haesbroeck, "Principal components analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies," *Ph.D. Dissertation, Univ. of Washington*, vol. 87, 603-618 2000.

[91] C. Croux and A. Ruiz-Gazen, "Breakdown estimators for principal components: the projection-pursuit approach revisited," *Journal of Multivariate Analysis*, vol. 95, 206-226 2005.

[92] J. Bouvrie, T. Ezzat, and T. Poggio, "Localized spectro-temporal cepstral analysis of speech," *In the proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASP)*, pp. 4733–4736, March 31-April 1 2008.

[93] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Winter 1991.

[94] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "Ecg analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, June 2001.

[95] J. Murakami, S.-i. Ito, Y. Mitsukura, and M. Jianting Cao; Fukumi, "A design of the eeg feature detection and condition classification," *In the proceedings of the Annual Conference SICE*, p. 27982803, September 2007.

[96] T. Lagerlund, F. Sharbrough, and N. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition," *Journal of Clinical Neurophysiol*, vol. 14, p. 73 83, 1997.

[97] A. Soong and Z. Koles, "Principal-component localization of the sources of the background eeg," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 59–67, January 1995.

[98] A. Hyvrinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, 411-430 2000.

[99] I. Navarro, F. Sepulveda, and B. Hubais, "A comparison of time, frequency and ica based features and five classifiers for wrist movement classification in eeg signals," *In the Proceedings of the IEEE International Conference of Engineering in Medicine and Biology Society*, vol. 4, pp. 2118–2121, March 2005.

[100] G. Herrero, A. Gotchev, I. Christov, and K. Egiazarian, "Feature extraction for heartbeat classification using independent component analysis and matching pursuits herrero," *In the Proceedings of the In the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 725–728, March 2005.

[101] M. Casey, "Reduced-rank spectra and minimum-entropy priors as consistent and reliable cues for generalized sound recognition," *in Proc. Workshop on Consisrenr and Reliable Acoustic Cues for Sound Analysis, Eumspeech, Aalborg, Denmark*, 2001.

[102] L. C.-T. Lin, W.-H. Chao, and S.-F. L. Yu-Chieh Chen, "Adaptive feature extractions in an eeg-based alertness estimation system," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2096–2101, October 2005.

[103] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics 5*, pp. 111–126, 1994.

[104] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature 401 (6755)*, 788-791 1999.

[105] ——, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, 556-562 2001.

[106] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756–2779, 2007.

[107] M. Berry, M. Browne, A. Langville, V. Pauca, and R. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, September 2007.

[108] V. Donoho, D.; Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" *available at http://www-stat.stanford.edu/ donoho*, 2003.

[109] D. Donoho and V. Stodden, "Seeding non-negative matrix factorizations with the spherical k-means clustering," *M.S. Thesis, University of Colorado*, 2003.

[110] C. Boutsidis and E. Gallopoulos, "Svd based initialization: A head start for nonnegative matrix factorization," *Pattern Recogn.*, vol. 41, no. 4, pp. 1350–1362, 2008.

[111] M. Carey, E. Parris, and H. Lloyd-Thomas, "A comparison of features for speech, music discrimination," *In the Proceedings of the International Conference Acoustics, Speech, and Signal Processing, Phoenix, AZ, USA*, vol. 1, pp. 149–152, March 1999.

[112] J. Saunders, "Real-time discrimination of broadcast speech/music," *In Proc. of International Conference Acoustics, Speech, and Signal Processing (ICASSP), Atlanta, GA,*, vol. 2, pp. 993–996, May 1996.

[113] J. Pinquier, J.-L. Rouas, and R. A. Obrecht, "Robust speech / music classification in audio documents," *International Conference on Spoken Language Processing, Denver, USA*, vol. 3, no. 3, pp. 2005–2008, September 2002.

[114] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," *In Proc. of International Conference Acoustics, Speech, and Signal Processing (ICASSP), Munich, Germany*, vol. 2, pp. 1331–1334, April 1997.

[115] N. Mesgarani, M. Slaney, and S. Shamma, "Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations," *In the Proceedings of the International Conference Acoustics, Speech, and Signal Processing, Phoenix, AZ, USA*, vol. 14, no. 3, pp. 920 – 930, May 2006.

[116] J. M. Kates, "Classification of background noises for hearing-aid applications," *The Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 461–470, January 1995.

[117] H. Deshpande, R. Singh, and U. Nam, "Classification of music signals in the visual domain," *Proc. the COSTG6 Conf. on Digital Audio Effects*, 2001.

[118] I. Paraskevas and E. Chilton, "Audio classification using acoustic images for retrieval from multimedia databases," *EURASIP Conference focused on Video/Image Processing and Multimedia Communications*, vol. 1, pp. 187 – 192, July 2003.

[119] S. Esmaili, S. Krishnan, and K. Raahemifar, "Content based audio classification and retrieval using joint time-frequency analysis," *In Proc. of International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 665–668, May 2004.

[120] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmental sound recognition using mp-based features," *In the Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008*, pp. 1 – 4, March 31 - April 4 2008.

[121] K. Umapathy, S. Krishnan, and S. Jimaa, "Multigroup classification of audio signals using time-frequency parameters," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 308 – 315, April 2005.

[122] A. Abu-El-Quran, R. Goubran, and A. Chan, "Adaptive feature selection for speech / music classification," *In the Proceedings of the IEEE 8th Workshop on Multimedia Signal Processing*, pp. 212–216, October 2006.

[123] L. Lu, H.-J. Zhang, and H. Jiang, "Content analysis for audio classification and segmenta-

tion," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 7, pp. 504 – 516, October 2002.

[124] G. Guodong and S. Li, "Content-based audio classification and retrieval by support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 209 – 215, January 2003.

[125] K. Umapathy, S. Krishnan, and R. Rao, "Audio signal feature extraction and classification using local discriminant bases," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1236 – 1246, May 2007.

[126] S. Xi, X. Changsheng, and M. Kankanhalli, "Unsupervised classification of music genre using hidden markov model," *In the proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, vol. 3, pp. 2023– 2026, June 2004.

[127] G. Freeman, R. Dony, and S. Areibi, "Audio environment classication for hearing aids using artificial neural networks with windowed input," *In the proceedings of the IEEE Symposium on Computational Intelligence in Image and Signal Processing*, vol. 2846, pp. 183 – 188, April 2007.

[128] M. Buchler, S. Allegro, S. Launer, and N. Dillier, "Sound classification in hearing aids inspired by auditory scene analysis," *EURASIP Journal on Applied Signal Processing*, no. 18, pp. 2991 – 3002, March 31 - April 4 2005.

[129] C. Panagiotakis and G. Tziritas, "A speech/music discriminator based on rms and zero-crossings," *IEEE Transactions on Multimedia*, vol. 1, no. 7, pp. 155 – 166, Febuary 2005.

[130] S. R. Gunn, "Support vector machines for classification and regression," *Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton*, 1997.

[131] T. Klingenheben, P. Ptaszynski, and S. Hohnloser, "Quantitative assessment of microvolt t-wave alternans in patients with congestive heart failure," *J. Cardiovasc Electrophysiol.*, vol. 16, pp. 620 – 624, 2005.

[132] D. M. Bloomfield, R. C. Steinman, P. B. Namerow, M. Parides, J. Davidenko, E. S. Kaufman, T. Shinn, A. Curtis, J. Fontaine, D. Holmes, A. Russo, C. Tang, and J. T. Bigger, "Microvolt t-wave alternans distinguishes between patients likely and patients not likely to benefit from implanted cardiac defibrillator therapy," *Circulation*, vol. 110, no. 14, pp. 1885–1889, October 2004.

[133] D. Judd, P. McKinley, and A. Jain, "Large-scale parallel data clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 876–876, 1998.

[134] S. Bhatia and J. Deogun, "Conceptual clustering in information retrieval," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 28, no. 3, pp. 427–436, 1998.

[135] C. Carpineto and G. Romano, "A lattice conceptual clustering system and its application to browsing retrieval," *Machine Learning*, vol. 24, no. 2, pp. 95–122, 1996.

[136] H. Frigui and R. Krishnapuram, "A robust competitive clustering algorithm with applications incomputer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 450–465, 1999.

[137] H. Abbas and M. Fahmy, "Neural networks for maximum likelihood clustering," *Signal Processing*, vol. 36, no. 1, pp. 111–126, 1994.

[138] R. Duda, P. Hart, and D. Stork, "Pattern classification," *Wiley New York*, 2001.

[139] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1928.

[140] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[141] M. Kyan, "Unsupervised learning through dynamic self-organization: Implications for microbiological image analysis," *In PhD thesis, School of Electrical and Information Engineering University of Sydney*, 2007.

[142] H. Kong and L. Guan, "Detection and removal of impulse noise by a neural network guided adaptive median filter," *In the proceedings of the IEEE International Conference on Neural Networks*, vol. 2, 1995.

[143] R. Sataloff, "Professional voice: the science and art of clinical care," *New York: Raven Press*.

[144] P. Carding and A. Wade, "Managing dysphonia caused by misuse and abuse," *BMJ 321*, p. 15445, 2000.

[145] E. Wallen and J. Hansen, "A screening test for speech pathology assessment using objective quality measures," *In the Proceedings of the International Conference on Spoken Language Proceedings, Philadelphia, Pa, USA*, vol. 2, pp. 776–779, October 1996.

[146] R. Moran, R. Reilly, P. de Chazal, and P. Lacy, "Telephony-based voice pathology assessment using automated speech analysis," *IEEE Transaction on Biomedical Engineering*, vol. 53, pp. 468–477, 2006.

[147] T. Ananthakrishna, K. Sharma, and U. Niranjan, "k-means nearest neighbor classifier for voice pathology," *In Proceedings of the IEEE India Conference INDICON, Indian Institute of Technology, Kharagpur*, pp. 352–354, 2004.

[148] A. Dibazar, S. Narayanan, and T. Berger, "Feature analysis for automatic detection of pathological speech," *In Proceedings of the EMBS Conference, Houston, USA*, pp. 182–183, 2002.

[149] J. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of shortterm cepstral parameters and neural network based detectors," *IEEE Transaction on Biomedical Engineering*, vol. 51, pp. 380–384, 2004.

[150] A. Gelzinis, A. Verikas, and M. Bacauskiene, "Automated speech analysis applied to laryngeal disease categorization," *Comput. Methods Prog. Biomed.*, vol. 91, no. 1, pp. 36–47, 2008.

[151] N. Senz-Lechna, J. I. Godino-Llorentea, V. Osma-Ruiza, and P. Gmez-Vildab, "Methodological issues in the development of automatic systems for voice pathology detection," *Elsevier, Biomedical Signal Processing and Control*, no. 11, pp. 120–128, 2006.

[152] V. Parsa and D. Jamieson, "Identification of pathological voices using glottal noise measures," *Journal Speech Language Hearing Research*, vol. 43, no. 2, pp. 469–485, 2000.

[153] K. Umapathy, S. Krishnan, V. Parsa, and D. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Transaction Biomedical Engineering*, vol. 52, pp. 421–430, 2005.

[154] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.

[155] S. Hadjitodorov and P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal disease screening," *Medical Engineering Physics*, vol. 24, no. 6, p. 419429, January 2002.

[156] C. Maguire, P. de Chazal, R. Reilly, and P. Lacy, "Identification of voice pathology using automated speech analysis," *in the Proceedings of MAVEBA*, p. 259262, December 2003.

[157] M. Marinaki, C. Kotropoulos, I. Pitas, and N. Maglaveras, "Automatic detection of vocal fold paralysis and edema," *in the Proceedings of INTERSPEECH - ICSLP*, pp. 537–540, November 2004.

[158] M. Wester, "Automatic classification of voice quality: comparing regression models and hidden markov models," *in the Proceedings of Voice data*, p. 9297, January 1998.