Theses and dissertations

1-1-2012

Signal Analysis In The Ambiguity Domain

Lakshmi Sugavaneswaran
Ryerson University

Signal Analysis in the Ambiguity Domain

by

Lakshmi Sugavaneswaran

B.Eng., University of Madras, India, June 2002

M.A.Sc., Concordia University, Montreal, Canada, November 2007

A dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2012

Author's Declaration

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

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

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

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

Abstract

SIGNAL ANALYSIS IN THE AMBIGUITY DOMAIN

©Lakshmi Sugavaneswaran, 2012

Doctor of Philosophy

in the Program of

Electrical and Computer Engineering

Ryerson University.

Time-Frequency Distributions (TFDs) are accounted to be one of the powerful tools for analysis of time-varying signals. Although a variety of TFDs have been proposed, most of their designs were targeted towards obtaining good visualization and limited work is available for characterization applications.

In this work, the characteristics of the ambiguity domain (AD) is suitably exploited to obtain a novel automated analysis scheme that preserves the inherent TF connection during Non-Stationary (NS) signal processing. Following this, an energy-based discriminative set of feature vectors for facilitating efficient characterization of the given time-varying input has been proposed. This scheme is motivated by the fact that, although, the interfering (or cross-) terms plague the representation, they carry important signal interaction information, which could be investigated for usability for time-varying signal analysis.

Once having assessed the suitability of this domain for NS signal analysis, a new formulation for obtaining AD transformation is introduced. The number theory concepts, specifically the even-ordered Ramanujan Sums (RS) are used to obtain the proposed transform function.

A detailed investigation and comparison to the classical approach, on this novel class of functions reveals the many benefits of the RS-modified AD functions: inherent sparsity in representation, dimensionality reduction, and robustness to noise.

The next contribution in this work, is the proposal of kernel modifications in AD for obtaining high resolution (and good time localization) distribution. This is motivated by the existing trade-off between TF resolution and interfering term reduction in TF distributions. Here, certain variants of TF kernels are proposed in the AD. In addition, kernels that are derived from the concept of learning machines are introduced for discriminative characterization of NS signals.

Following this, two novel AD-based schemes for neurological disorder discrimination using gait and pathological speech detection are introduced. The performance evaluation of these AD-based schemes, using a linear classifier, resulted in a maximum overall classification accuracy of 93.1% and 97.5% for gait and pathological speech applications respectively. The accuracies were obtained after a rigorous leave-one-out technique validation strategy. These results further confirm the potential of the proposed schemes for efficient information extraction for real-life signals.

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible.

First and foremost, I would like to express my sincere gratitude to my mentors Professor. Sridhar Krishnan and Dr. Karthikeyan Umapathy for being a constant source of encouragement and for their thoughtful guidance throughout this journey. The enthusiasm they shared for research has been extremely contagious and motivational, especially during the tough times when I seemed lost in the PhD pursuit. If not for their contributions of motivation, inputs, ideas and funding, this would have been an extremely challenging venture.

I am also grateful to Dr. Boris Steipe, Dr. John Marshall, Dr. Marie Killeen, Dr. Yifeng He and to all the wonderful professors from Ryerson, I had an opportunity to learn from and to interact. My special thanks to my PhD committee members: Dr. James Smith, Dr. Dimitri Androutsos, Dr. Aziz Guergachi and Dr. Ramin Pichevar, for their constructive feedbacks during different stages of this dissertation. I would like to take this moment to thank all my friends and colleagues from SAR lab, for their helpful feedback during each stage of this work.

Lastly, and most importantly, I wish to thank my family and friends for their unconditional love and encouragement. I am ever indebted to my parents and brother for raising me with a love of science and for supporting my pursuits. To them I dedicate this dissertation!

Contents

List of Tables

List of Figures

List of Acronyms

AD	Ambiguity Domain
AF	Ambiguity Function
ALS	Amytrophic Lateral Sclerosis
CART	Classification And Regression Tree
CSK	Compact Support Kernel
CWD	Choi-Williams Distribution
DFT	Discrete Fourier Transform
ECG	Electrocardiogram
EMG	Electromyography
FFT	Fast Fourier Transform
FM	Frequency Marginals
FS	Frequency Support
GCD	Greatest Common Divisor
HD	Huntington's Disease
ICA	Independent Component Analysis
IF	Instantaneous Frequency
IQR	Inter-Quartile Range
LDA	Linear Discriminant Analysis
MEEI	Massachusetts Eye and Ear Infirmary
MHD	Margenau-Hill Distribution
ML	Machine Learning
NMF	Non-negative Matrix Factorization
NN	Neural Network
NS	Non-Stationary

TF	Time-Frequency
TFD	Time-Frequency Distribution
PCA	Principal Component Analysis
PD	Parkinson's Disease
RDR	Relative Distribution Ratio
RE	Realness
RFT	Ramanujan Fourier Transform
RID	Reduced Interference Distribution
RS	Ramanujan Sums
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TD	Time Delay
TFR	Time-Frequency Representation
TM	Time Marginals
TS	Time Support
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

Chapter 1

Introduction

1.1 Motivation

The pursuit of many researchers, in the field of information processing, has been to develop thinking machines that could mimic the human brain's intelligence and solve problems [2–4]. Design of such tools requires a reasonable understanding of the underlying behavior of the governing system. In the past century, a variety of investigations [3,5] were carried out from a signal analysis perspective to unravel the underlying information in a given system. For example, use of change in signal characteristics for identifying the warning signals in financial markets, and early detection of human system altercations.

In the past, these analyses were carried out manually, depending heavily on the investigators' expertise in that field. Such approaches, although usable, are becoming less suitable when the amount of data increases. Hence, there exists an increased interest for automated signal analysis schemes. The advancement in the signal collection strategies extends the applicability of signal processing to a variety of fields, such as, biosignals, bioinformatics, communication systems, geoinformatics, quantum physics and security.

Automatic signal processing tools provides the following distinct advantages over manual

analysis methods:

- Reliable: Reliable accuracy characteristics and performance independent on the total amount of data to be processed.

- Computationally Efficient: Ease of performing advanced computations.

- Objective Assessment: Provides quantitative metrics which are purely objective and hence, is not affected by the individuals state-of-mind like manual analysis techniques.

- Ease of Detection: Can facilitate detection of critical transition regions for long signals, which can otherwise be missed or overlooked in manual processing.

- Long-term monitoring: Use of automated techniques can facilitate analysis of long signals for implementation of screening units for patients requiring long-term care.

- Development of new techniques: Owing to the re-usability, these systems can be extended for obtaining more optimized tools for different kinds of applications.

Considering the many advantages of automated signal processing, the main focus of this dissertation is to develop robust and efficient signal processing techniques for analyzing complex dynamic systems with applications to biomedical signal characterization.

1.2 Signal Categories

A signal can be defined as a function that carries information about the attributes (both significant and redundant) of its generating source. Mathematically, any quantity that exhibits a variation in time or frequency or space is potentially a signal. The most common distinction among signals depends on whether the signal is defined in continuous or discrete time intervals. Owing to the complexity associated with continuous time processing and during realization, most computations are carried out in the discrete-time domain. From a
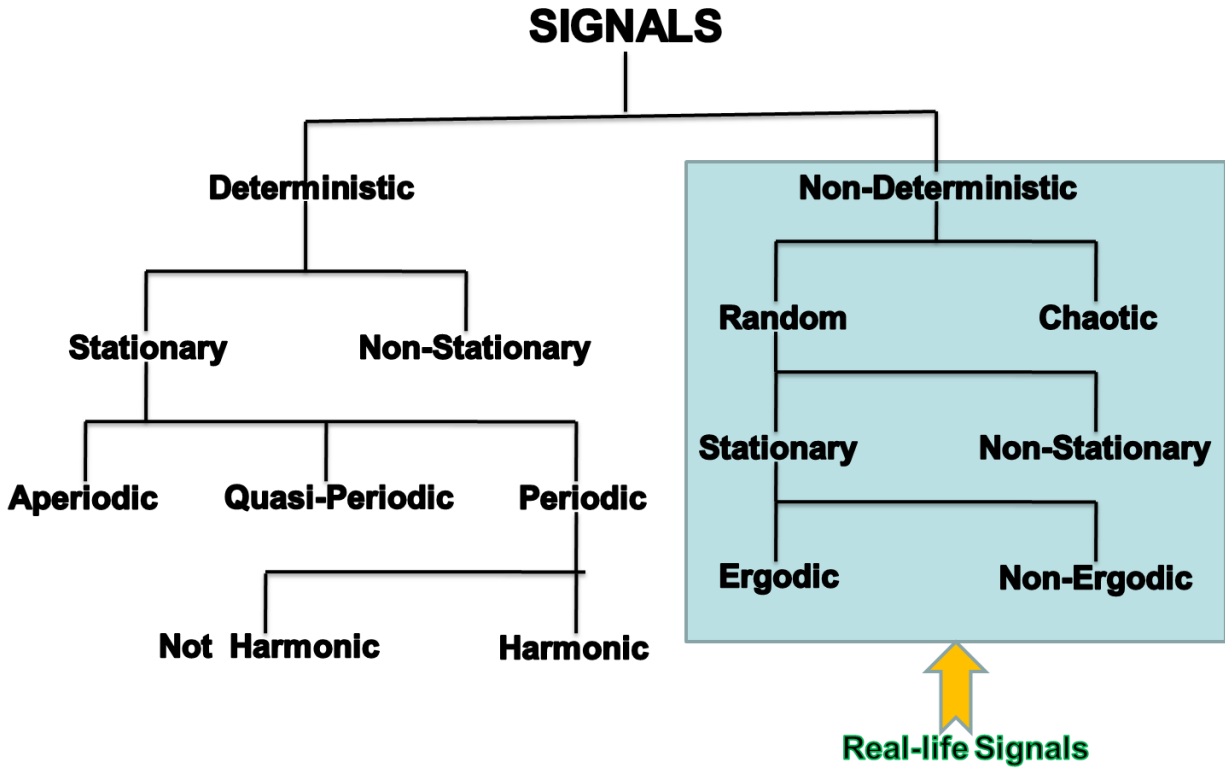
Figure 1.1: General Classification of Signals

signal processing sense, the continuous and discrete signals can be sub-divided [6] into two types: deterministic and non-deterministic.

A signal is said to be deterministic, if the analytical expressions for past, present and future instants of time can be described and the signal can be predicted and reproduced. Periodic signals (eg: sinusoid) and periodic extensions are some examples of deterministic signals. In simple terms, all deterministic signals have modeling equations that can be used to predict the behavior at any instant of time. For a given periodic signal with period T, the following relation holds true.

$$x(t) = x(t + T) \cdot \tag{1.1}$$

Non-deterministic signals [7] are those signals, whose behavior cannot be determined and

3
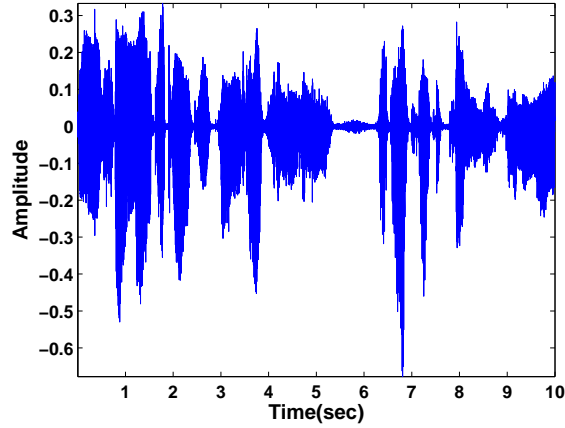
Figure 1.2: Pathological Speech Time-Series Data

for which there are no explicit mathematical models. Such signals usually take random values at any given time. These signals can further be divided into stationary and Non-Stationary (NS) signals. The outcomes of an unbiased die are an example of a stationary non-deterministic signal. The signal from a die is considered stationary because, the number of faces is fixed and the occurrence of them are equiprobable (1/6).

Figure 1.1 shows the overall classification of the signal categories. Most real-world signals are non-deterministic in nature, due to the non-availability of suitable models. In addition, these signals are mostly NS since the inherent system dynamics are time-varying in nature. Some of the instances for such systems are uncertainty associated with the weather leading to climatic bifurcations, abrupt transitions in ecosystem, stock market fluctuations and biomedical systems. Figure 1.2 shows an example of a 10 second pathological speech signal [8].

From the figure, it can be observed that such signals have varying mean and variance measures at different instants of time, and hence, carry non-stationary behavior. However, since they carry significant information about the underlying phenomenon that generated them, their analysis needs to be carried out properly and a suitable representation technique

4

needs to be selected for efficient characterization. The design of these schemes can assist in facilitating diagnosis (or screening) and thereby, can help improve the overall quality of life. Hence, there exists a growing need for automated signal analysis techniques for processing real-world signals. In the following sub-section, the choice of the most suitable signal representation tool will be investigated.

1.3   Automated Non-Stationary Signal Analysis

Most real-world signals are NS in nature and they carry information about the generating phenomenon. In order to build an automated signal analysis scheme, three fundamental modules are required: Signal Representation, Feature Extraction, and Classification. These modules will be discussed in detail in the following sub-sections.

1.3.1   Representation Tools

Owing to the inherent non-deterministic nature of NS signals, their analysis tends to be more complex and challenging. In order to develop efficient analysis schemes, the signal first needs to be represented in an appropriate domain, where the inherent signal behavior can be better understood. Three main representation domains are available for this purpose:

Time: One of the standard ways by which a signal is represented is in time domain. The time-domain graph shows the changes in signal behavior for different instants of time. In order to obtain an understanding of each domain, a bird sound example is used. This signal represents a simple multicomponent time-varying signal example. From Figure 1.3 (a), it can be observed that there exists certain repetitive patterns in the sound. The commonly extracted metrics in time domain consists of first order and second order statistics, such as, amplitude range, average, variance measure, and correlation.

Frequency: A time-series data can be represented in a variety of different ways depending

on specific application. In comparison to the time domain, the significant features of a given signal can be better characterized in the frequency domain. The time domain defines the change in the signals' amplitude as a function of time, while in the frequency domain the rate at which the signal amplitude changes at different instants of time can be better quantified. The conventional Fourier analysis is one of the commonly used techniques to transform the time series into spectral domain and to allow computation of the rate of change of signal characteristics over time. Here, a given signal is decomposed into sum of weighted sinusoidal functions and their relative intensity is computed. Any given signal can be mapped from time domain to the frequency domain and vice-versa, using transformation functions (such as Fourier). The spectral domain provides an energy distribution map for different frequencies, as can be verified from Figure 1.3 (b).

Time-Frequency: Time- and Frequency domain shows the signal behavior with reference to a single variable change (either time or frequency). To get a better understanding of the signal dynamics for multicomponent analysis, techniques that allow simultaneous variation of both time and frequency variables needs to be used. One such method is joint Time-Frequency (TF) analysis. As can be noted from the Figure 1.3(c), this plane shows the energy of the signal for different TF values.

This domain provides meaningful information about the entire signal by preserving the inherent TF connection of the system. That is, information regarding time-varying frequency distribution can be distinctly quantified in comparison to time and frequency domains.

Choice of TF Analysis

As discussed in the previous section, for the case of the bird sound example, the time domain provided a good observation medium for time-sample repeats while the spectral plot revealed the frequency distribution information. To understand the overall behavior of the signal, the TF representation revealed the signal trend along with details regarding the instantaneous
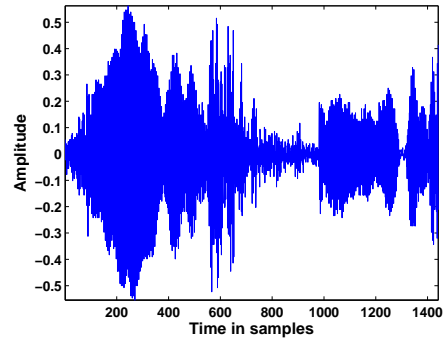
(a)
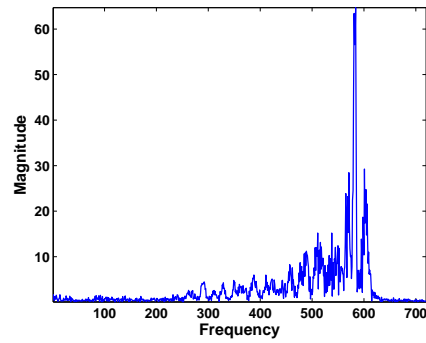
(b)
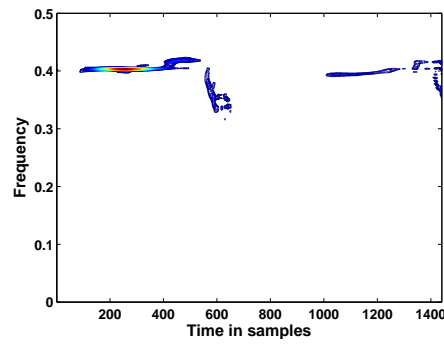
(c)

Figure 1.3: (a) the bird sound time-series plot. (b) Spectral representation of the bird sound. (c) Time-Frequency map of the bird sound.

energy distribution over time and frequency.

Since real-life signals have time-varying behavior and the frequency content evolves over time, the choice of representation should accommodate both instantaneous energy computation and preserve the behavior trend.

In order to visually understand this, let us consider an example of chirps (see Figure 1.4): increasing and decreasing. In time domain, the signal frequency change information is not distinctly available. Further, in frequency domain the representation remains the same for both increasing and decreasing frequency trends, indicating the fact the Fourier representation fails when the signal carried time-varying behavior. However, TF plane not only succeeds in capturing the frequency distribution at various time instants but also provides the phase trend information. This provides the validation of the suitability of this plane for design of NS signal analysis systems.

1.3.2 Feature Extraction

For complex signal analysis, the signals are first represented in a suitable domain where the underlying characteristics are preserved and following this, the input is transformed into a set of features. The number of features is usually much less than the overall dimension of the input and/or the representation domain dimension. The choice of features usually depends on the specifications of the desired task. Feature extraction is a general term used for obtaining an approximate set of signal attributes for describing a given signal and is considered to be a means for achieving dimensionality reduction.

Since no apriori information is available for most real world signals, extraction of discriminative features can facilitate efficient characterization of a given NS signal. In general, the extracted features [3] need to satisfy the following criteria:
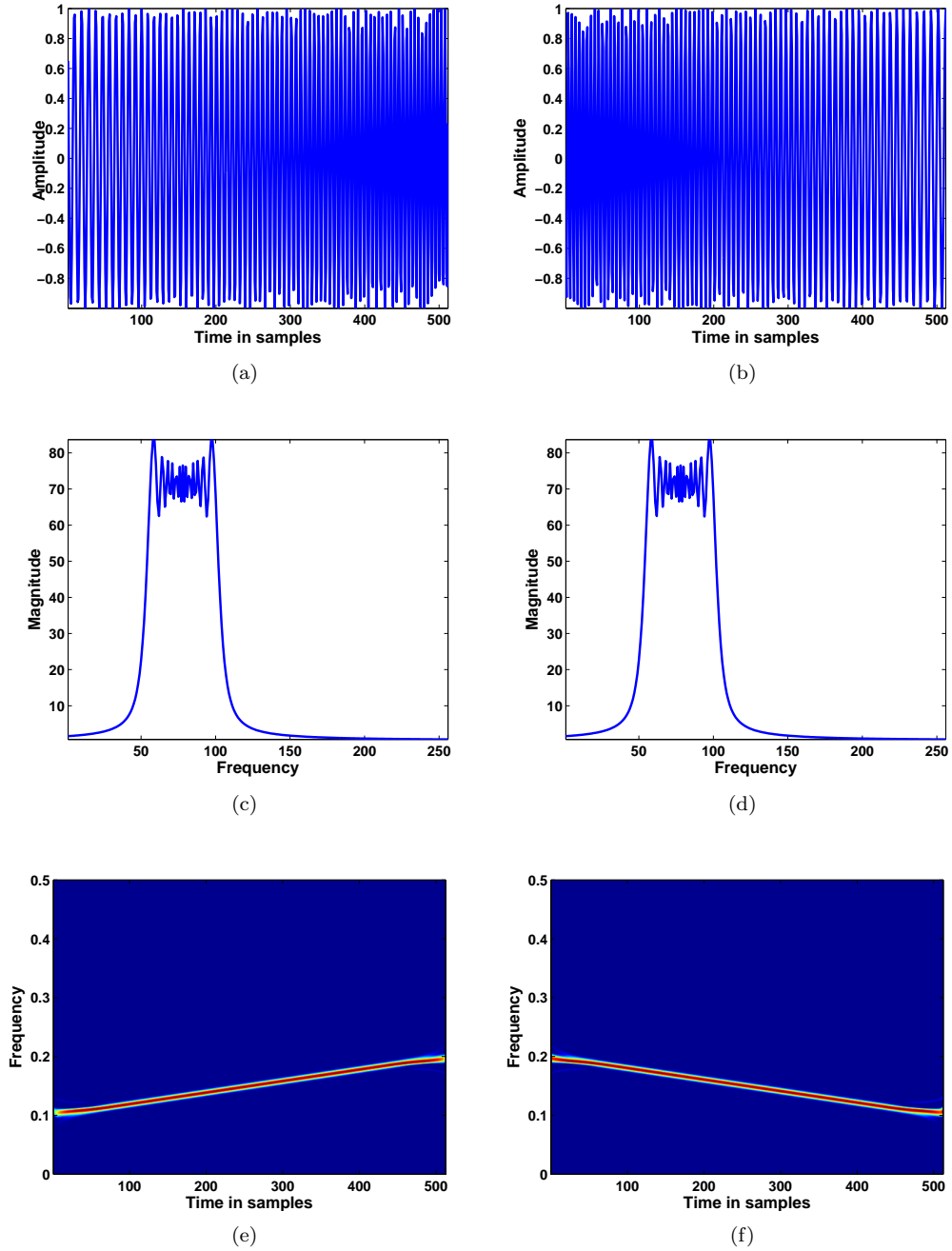
Figure 1.4: (a) Increasing Chirp. (b) Decreasing Chirp. (c) and (d) Frequency representation of increasing and decreasing chirp respectively. (e) TF representation of increasing chirp. (f) TF representation of decreasing chirp.

9

- Discriminative: the features should be able to characterize the signal properties and be able to uniquely define different classes of signals

- Representative: should carry relevant information from the signal with respect to the application in order to facilitate reduced dimension processing

- Robust: The extracted features should carry meaningful information about discrimination in TF pattern and be robust to time and frequency changes between samples within a given class, to accommodate natural uncertainties when handling real-life signals.

1.3.3 Classification

The third important block of an automated signal analysis unit is the classification module. Classification refers to the process of formulating prediction rules for a given set of observations and their corresponding class information. A classification scheme employs two phases: training and testing. In the training phase, a self-learning algorithm with integrated optimization (cost function to control the number of iterations) mechanism trains the classifier for achieving maximum performance metrics. The trained system is then validated using different combinations of the given set of test cases. The overall accuracy is computed as the average of all test case combination performances.

In order to robustly assess the effectiveness of an approach and to obtain a rigorous evaluation of the feature space, a classifier that uses a robust cross-validation approach needs to be chosen. In this work, the proposed approaches are rigorously validation using the linear discriminant analysis technique [5].

1.4 Previous Studies in Non-Stationary Signal Analysis

The classical approach in NS signal processing is to obtain a reasonably small interval of data by segmentation. This segmented data is considered stationary and is then processed using available stationary signal processing techniques. The main limitation associated with this technique is the assumption, stating a random signal as being stationary in fixed time sections, is loosely framed since most real signals contain time-varying behavior throughout. In other words, the segmented sections may exhibit non-stationarities, even within small observation windows where it is expected to be stable. The processing of such signals have been addressed using a variety of representation schemes.

Following the classical approach, the first set of proposed methods used non-parametric TF representations such as the Short-Time Fourier Transform (STFT) and spectrogram for estimation purposes. Further probing into the signal characteristics resulted in second set of approaches that extracts relevant features were employed for speech signal applications. These methods were developed using Hilbert transform [9] or by applying energy operators like the Teager-Kaiser operator [10]. The study shows that using the Hilbert transform for peak TF plots shows large false frequency indications when the time series model of the input undergoes abrupt frequency changes. Also, when multiple dominant frequencies are present in the signal, corresponding Hilbert transform plots do not always provide clear results owing to the effects of low-frequency terms in the input. For the case of the energy operator, there is still room for enhancement mainly because of the fact that presence of noise during multi-component analysis greatly affects the performance of the Teager operator and higher order models are sought after for achieving little improvement in operation efficiency. In short, the limitation of these schemes is the lack of robustness, especially for multi-component signal analysis.

An extension to the "best-basis" method for classification applications was proposed by Saito

et al [11]. This method addressed the issue related to reducing the dimensionality of the input signal or image for processing using the traditional Linear Discriminant Analysis (LDA) classifier and Classification and Regression Tree (CART) classifier. Various other filtering-based schemes followed, whereby preprocessing techniques were enhanced for efficient removal of visual artifacts to facilitate better classification. Some of these schemes employed modified Kalman filtering technique. These schemes [4, 12] were tested for the electrocardiogram (ECG), Polysomnography (PSG) signals. Several methods have been proposed to classify biomedical NS signals and features such as temporal intervals, spectral features and the transform coefficients have been widely analyzed. Some of the popular classification techniques include linear discriminant analysis, Bayesian methods, support vector machines, independent component analysis and artificial neural networks. For instance, Jiang et al [13] investigated the application of block-based neural network for ECG classification. Study shows that these methods reinforce the non-stationarity concept and call for the necessity to extract more relevant information to study the dynamics of the signal under study.

For detecting and analyzing NS signals, TF Distribution (TFD) of energy has been recognized as a powerful tool. The TFDs allow efficient visualization of the evolving frequency behavior during some NS events by providing a reasonable map of the time signal in the TF plane. Cohen's class of TFD [14] is one of the widely used techniques and is viewed as the Fourier transform of a weighted Ambiguity Function (AF). This formulation led other researchers to introduce the concept of "kernel", and also provided an important model to obtain many other types of TFDs for a variety of NS signal analysis applications. The choice of kernels further determined the efficiency of a TFD in regards to a specific input pattern. That is, the use of kernel functions directly affects time resolution, frequency resolution and conformity to marginal conditions. Additional details on available kernel designs and their effect on TFD analysis will be provided in the following sub-section and in Chapter 2.

1.5   Contributions

This work presents a generalized ambiguity domain (AD) based methodology that exploits the benefits of the AD for achieving automated signal analysis and characterization of NS signals. An overview of the proposed framework and the contribution modules (highlighted) are shown in Figure 1.5.

Signal Analysis using ambiguity domain:

The main aim of this dissertation is to provide a novel analysis scheme that preserves the inherent TF connection in NS signals. To achieve this objective, the first contribution will target design of a reasonably high-performing analysis approach by exploiting the ambiguity domain (AD), the non-negative representation of the Wigner-Ville distribution. The novelty of the approach lies in the fact that the interfering terms generated during bilinear transformation, which are otherwise removed to improve visual clarity, is used to analyze a given NS signal. This is motivated by the fact that, although, they plague the representation, they carry important signal interaction information, which could be investigated for usability for time-varying signal analysis. After having obtained the desired scheme, a discriminative set of feature vectors defined by energy values in the AD-plane, is extracted from the transformed signal. Finally, a detailed investigation on the signal interaction terms is presented.

Transformation :

Once the AD-based analysis scheme is obtained, a novel AD transformation based on number theory is introduced. The proposed transform aims at introducing inherent sparsity to the TF functions and thereby obtaining good dimensionality reduction while processing certain kinds of signals. The proposed approach will be evaluated by comparing with the classical approach. Finally, this technique will then be extended for facilitation of TF analysis for NS signals.

Kernel Design in AD:
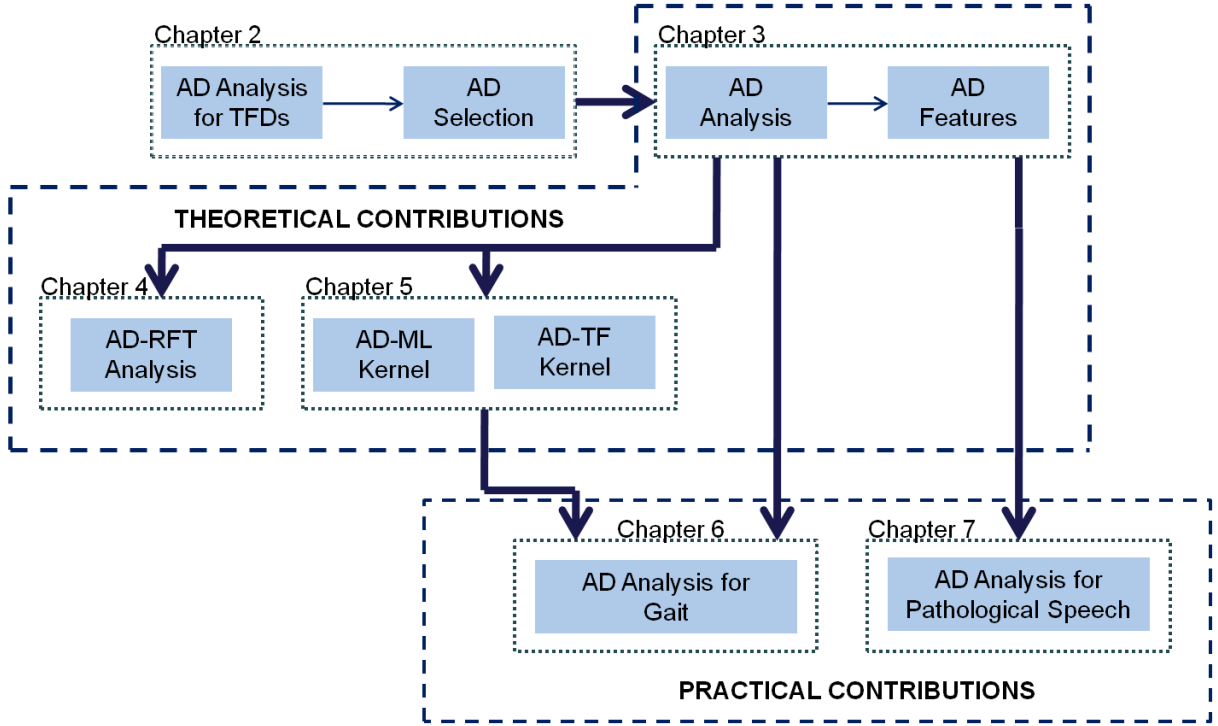
Figure 1.5: Block diagram of the Dissertation

The next contribution in this dissertation is motivated by the existing trade-off between TF resolution and interfering term reduction in TF distributions. To overcome this, a TF kernel modification and a machine learning based kernel is proposed in AD. All the main contributions are validated suitably using synthetic and some real-life signals.

1.6   Dissertation Structure

The dissertation is organized in 8 chapters:

Chapter 2: Non-Stationary Signal Analysis

This chapter will provide a concise overview about some of the existing linear and quadratic TF distributions and related previous work. Synthetic examples to demonstrate the properties of the different distributions and their usability for NS signal analysis will be investigated

to select a suitable representation domain.

Chapter 3: Signal Analysis in AD

A detailed investigation on the AD for signal analysis will be introduced in this chapter. In addition, the usability of this domain for signal characterization and their resemblance in the TF space will be studied. Further, a novel approach for the feature extraction using AD coefficients will be presented.

Chapter 4: AD Transformation using Ramanujan Sums

The main objective of this chapter is to introduce a novel class of Ramanujan Fourier Transform (RFT) based TF transform functions, constituted by Ramanujan sums (RS) basis, for analysis and estimation of signals. In addition, this work also provides a 2-D formulation of TF-RFT function.

Chapter 5: Kernel Design in AD

The objective of this chapter is to introduce modifications to TF kernels for obtaining efficient signal characterization. In addition, the usability of support vector machines (SVM) kernels for TF analysis were investigated by incorporating them in the AD. Following the proposal of the three schemes that analyze AD for signal characterization, Chapters 6 and 7 describe the practical contributions of these in this dissertation. Here, the usability of this domain to characterize certain real-life signals that show some inherent "rhythmic" behavior is discussed. The analyzed real-world signals were obtained from public data archives and the performance of the proposed schemes were benchmarked with state-of-the-art algorithms.

Chapter 6: AD Analysis of Gait Pattern

In Chapter 6, a robust AD-based characterization scheme that incorporates a novel feature extraction algorithm for screening the presence of a neurological disorder from normal gait data has been proposed.

Chapter 7: AD Analysis of Pathological Speech

Chapter 7 presents a novel approach for the analysis and classification of normal and patho-

logical speech signals. A detailed validation is carried out using the MEEI Voice disorders database.

Chapter 8: Conclusion and Future Work

Here, a summary of the complete dissertation is presented and the main contributions are highlighted. The novelty and the benefits of the proposed strategies are validated using certain synthetic and real-life examples. The potential of the proposed work for future research is introduced.

Chapter 2

Non-Stationary Signal Analysis

Real-life signals are NS in nature and therefore, their characterization necessitates the use of suitable transform functions that preserve the inherent TF connection in a given signal. Analysis of Discrete Fourier Transform (DFT) and time-domain schemes indicate the limitations of these techniques for analyzing signals that carry time-varying behavior. To overcome the shortcomings of the conventional techniques, joint TF analysis are often preferred for time-varying signal analysis.Although literature [14] shows the existence of a variety of TFDs for different applications, a concise overview of the available methods is presented in the following sections, in an effort to justify the selection of TFD in this research.

Depending on the structure of the transform function, TFDs can be broadly categorized into two forms: linear and quadratic. The difference is due to the order of the transform function and the presence/absence of interfering signals in the representation. A given signal, when mapped on a higher dimensional space, will possess the signal terms (or auto-terms) and may also have some generated interfering terms (or cross-terms) depending on the type of transformation. This chapter provides the rationale behind the choice of transformation function (as highlighted in Figure 2.1) and in the following sub-sections, some of the existing forms of TFDs will be discussed.
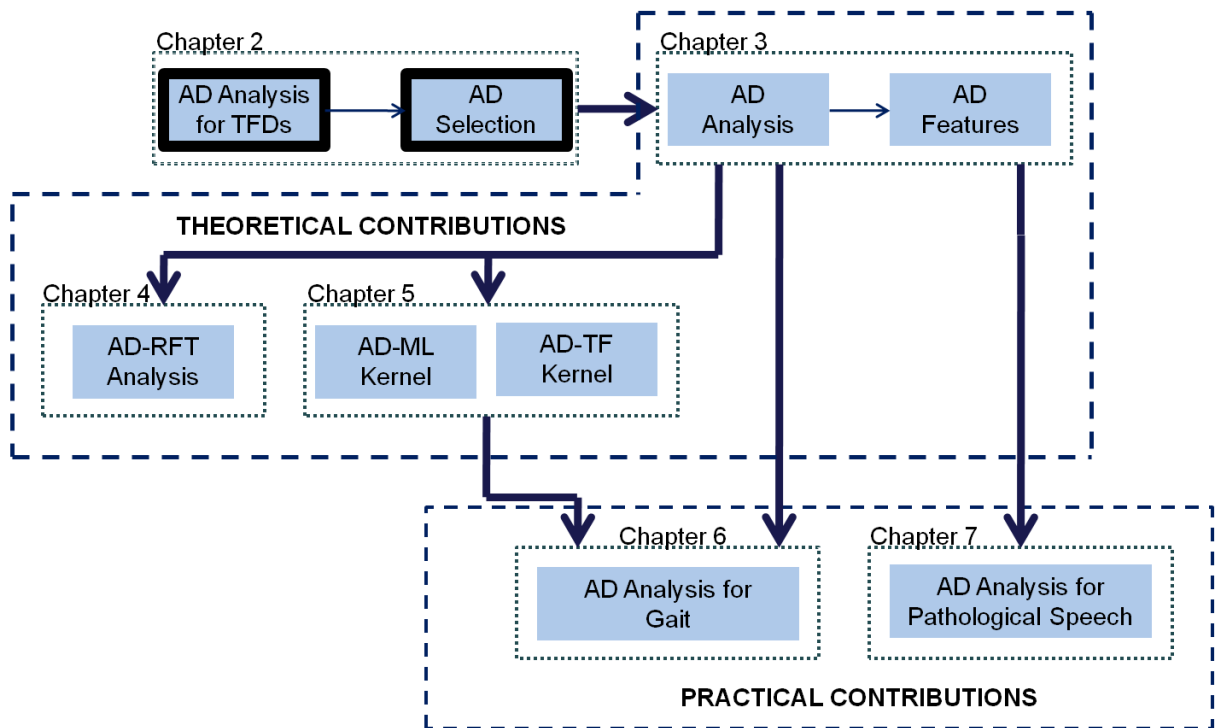
Figure 2.1: Chapter 2 - Non-Stationary Signal Analysis

2.1 Linear TFDs

For any given signal $x(t)$, a linear TFD is calculated as a function of the first order representation of the signal itself. A linear TFD obeys the superposition principle for a linear combination of signals. Let $L_{x_1}(t, f)$ and $L_{x_2}(t, f)$ be the linear TFDs of the signals $x_1(t)$ and $x_2(t)$ respectively. Then,

$$L_{ax_1+bx_2}(t, f) = aL_{x_1}(t, f) + bL_{x_2}(t, f). \tag{2.1}$$

A linear TF transformation of multicomponent signals generates only the signal terms, owing to the absence of any higher order frequency interactions and hence, no interfering terms. There exist three widely used classes of Linear TFDs: the STFT, the Wavelet Transform (WT) and the Gabor transform.

2.1.1 Short-Time Fourier Transform and Spectrogram

The STFT is often referred to as the "windowed Fourier transform" (see Figure 2.2) and for a given sliding window function, $\gamma^*(\tau)$, is defined as follows,

$$P_{STFT}(t, f) = \int_\tau x(\tau)\gamma^*(\tau - t)e^{-j2\pi f\tau}d\tau. \tag{2.2}$$

For analysis of transients only the real value of this representation i.e. the squared magnitude of P_{STFT} is used. This real formulation of STFT is usually referred to as the spectrogram, which is a quadratic TFD. The spectrogram representation shows the frequency distribution information for each time segment (determined by the window).
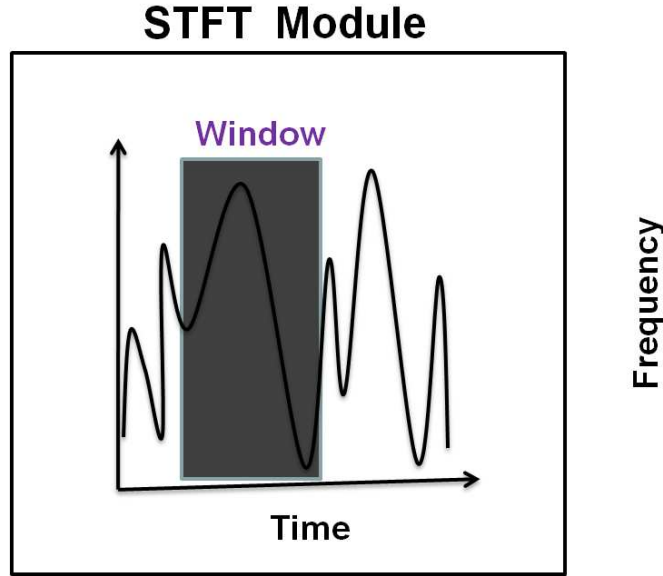
Figure 2.2: Representation of STFT Analysis

2.1.2 Gabor Transform

The Gabor transform is a special case of STFT defined using a Gaussian window function and is represented as,

$$P_{GT}(t, f) = \int_{-\infty}^{\infty} x(\tau)e^{-\pi(\tau-t)^2}e^{-j2\pi f\tau}d\tau. \tag{2.3}$$

Although the Gaussian function has an infinite range, for obtaining realizable discrete time processing applications, a significance level is chosen in order to limit the distribution. The bandwidth limited Gabor transform is then used in the analysis.

2.1.3 Wavelet Transform and Scalogram

A wavelet function is characterized by wave-like brief oscillations whose amplitude start at zero, have periodic or aperiodic increases and finally decreases back to zero. Depending on the definition of the intermediate pattern (increase-decrease), different classes of mother

wavelets evolved. For a given mother wavelet $\gamma^*(\tau)$, the wavelet transform (WT) of a signal $x(t)$ is defined as,

$$P_{WT}(t, f) = \sqrt{\left|\frac{f}{f_0}\right|} \int_\tau x(\tau)\gamma^*\left(\frac{f}{f_0}(\tau - t)\right) d\tau. \tag{2.4}$$

Here, the frequency ratio $\frac{f}{f_0}$ denotes the scaling factor for the time-shifted mother wavelet. Since a frequency ratio is used, WT is more appropriately referred to as the time-scale representation.

The mother wavelet is scaled into different levels and is slid across the signal for localizing the signal structures efficiently at each scale i.e. the scaled wavelets are used as dynamic windows for obtaining reasonable TF characterization of the signal for all instants of time. The wavelet scalogram displays the TF structure obtained from the WT. The scalogram representation defines the time and scale for a given wavelet scale.

2.2 Quadratic TFDs

Although linear TFDs are computationally efficient and do not suffer from cross-terms, these advantages come at the expense of limited TF resolution and choice of window function (limiting the allowable spread in time and frequency). Quadratic TF analysis is obtained by using the multiplicative comparison of the signal with its time-shifted version. Because of the presence of second-order computation, the principle of superposition does not hold good for this class of transforms. Some of the commonly used quadratic TFDs are discussed in the following sections.

2.2.1 Wigner-Ville Distribution

Cohen [14] proposed a generalized framework for obtaining different quadratic TFD formulations using filter functions (or kernels), following which a variety of distributions were

developed. The choice of an all-pass filter corresponds to the widely used Wigner-Ville distribution (WVD).

$$P_W(t, f) = \int_\tau x \left(t + \frac{\tau}{2} \right) x^* \left(t - \frac{\tau}{2} \right) e^{-j2\pi f\tau} d\tau. \tag{2.5}$$

Being of a bilinear disposition, for multicomponent signal analysis in addition to signal terms, cross-terms also result in this class of distributions. WVD offers good TF localization, satisfies marginals, maximal auto-term concentration and most other desirable properties (as listed in Section 2.4.1). The quadratic class can be considered as the smoothed versions of WVD since kernel functions are used for filtering out the cross-terms.

2.2.2 Smoothed Distributions

The Choi-Williams distribution is one of the initial works on kernel modified Cohen's class of functions, designed to reduce the interfering signal issue in WVD and the window length limitation in STFT. It is defined as follows:

$$P_{CWD}(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \left(t + \frac{\tau}{2} \right) x^* \left(t - \frac{\tau}{2} \right) e^{-\alpha(\eta\tau)^2} e^{j2\pi(\eta t - \tau f)} dt \, d\eta \, d\tau. \tag{2.6}$$

An exponential kernel function is used in this distribution in order to suppress the cross-terms in WVD. The value of α is adjusted according to the application. Although CWD succeeds in suppressing cross-terms, it offers poor TF localization and the representation is smeared by low pass filtering (exponential kernel).

Following this, a variety of TF Representations (TFRs) (like Page distribution, Margenau-Hill distribution, Zhao-Atlas-Marks distribution, etc) were developed to obtain reasonable TF localization and maximum auto-term concentration. All these techniques perform some type of smoothing on the WVD. However, the resulting TFRs had smearing of signal terms, along with tradeoff between cross-term suppression and auto-term concentration.

2.2.3 Other Reduced Interference Distributions

Following to the definition of CWD, many other kernel formulations followed, for obtaining a reasonable tradeoff between cross-term suppression and TF resolution. The reduced interference distribution (RID) class of functions were one of the recent inclusions in this respect. They are defined as follows,

$$P_{RID}(t,f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{|\tau|} h\left(\frac{\eta - t}{\tau}\right) x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-\alpha(\eta\tau)^2} e^{j2\pi(\eta t - \tau f)} d\eta \, d\tau. \quad (2.7)$$

Jeong et al [15] proposed the RIDs in an effort to overcome the limitations with WVD: negative values leading to misinterpretations and cross-terms. This class of transforms, offer good time and frequency support in addition to cross-term reduction.

2.3 Review of the state-of-the-art in TFD design

In recent years, TF distributions have been extensively studied for NS signal analysis. Most of the existing distributions describe how the frequency content of a signal varies with time. The spectrogram, the most common tool for this purpose, provides non-negativity and smoothed cross terms in the resulting distribution. However, this still involves a significant amount of tradeoff between time and frequency resolution. In order to achieve high TF resolution recent advances sacrifice on the non-negativity and introduces other distributions.

The Wigner distribution, one of the best known TF distributions, suffers owing to the substantial presence of cross terms between frequency components in the TF plane [16]. Attempts to remove cross terms lead to violation of certain desirable properties such as marginal conditions (i.e. presence of correct time and frequency marginals). Choi and Williams [17] introduced a distribution that uses an exponential kernel for reducing cross-terms suffer from smearing in the representation. This distribution was still found to be less efficient with ref-

erence to use with biomedical signals. Jeong and Williams [15] came up with a new class of TF distributions called the Reduced Interference Distributions (RID) which also suffered the previously mentioned drawbacks in terms of achieving high resolution and good localization. Studies suggest that signal-dependent TF representations perform well for a wide range of signals in comparison to fixed-kernel distributions for applications related to NS signal processing. Guo et al [18] proposed a Bessel kernel based TF distribution and tested it on Doppler blood flow and heart sound signals. Here, the finite support properties of the kernel were sacrificed to obtain the desired distribution. A TF kernel design [19] formulated by Hearon et al is obtained by constraining the minimum variance. Although this kernel provides the global minimum value for variance, the results show that the performance degrades in the presence of Gaussian noise.

Costa et al [20], Amin et al [21] and Jemili et al [22] proposed certain TF representations using exponential kernels, modified comb filters and wavelet function respectively. These schemes aimed at obtaining better signal localization. Tagluk et al [23] designed a universal kernel that uses the cross-terms generated to use the cross terms generated by Wigner-Ville Distribution (WVD) for detecting low level potentials. Following these, studies [24–26] concentrated on TF-based analysis of NS signals using certain other kernels such as cone-shaped kernels, Compact Support Kernel (CSK) and figure eight kernel.

Jones et al [27] developed a signal-dependent radially Gaussian kernel that has been tuned to adapt well with time, to track signal component variations over time and this design supported online implementation for signals of arbitrary length. Their method uses a short-time AF computed from the constant-time slices of the TF representation and it aims at optimizing the kernel function. This technique was shown to relatively outperform certain fixed kernel representation at the expense of increased resources, thereby cost and reduced resolution.

In addition to TF kernels, certain time-efficient Support Vector Machine (SVM) based novel

kernel functions [28,29] were obtained from the reproducing kernels of Hilbert space for classification of biomedical signals. In order to achieve better classification accuracy, attempts to reduce the noise component by using a specific filter in the desired signal were made by Pander et al. The filter aimed at minimizing the distance measure in the higher feature space to reduce the residual distortion for the case of ECG signal.

One of the alternatives [30] used to reduce the cross-terms is design of kernels for TFDs using the phase characteristics, rather than the amplitude response. These phase kernels do not attenuate cross-components but they translate them in the TF plane. These transformed cross-components are placed on top of the auto-components, after which a "don't care" region is specified for these unwanted components. However, the tradeoff between achieving minimum noise component and maximum retention of signal still remains more pronounced for NS signal data.

Ghoraani et al [31] has been the primary work to have investigated the feature localization aspect of the TF representation, by transforming the signals into matrices by employing the positive TF transform, and other representation techniques such as principal component analysis (PCA), independent component analysis (ICA), and non-negative matrix factorization (NMF). Their results promise a robust outcome and also offer a very high localization accuracy of about 96% for noisy input conditions.

Pekalska et al [32] proposed extensions of kernel linear and quadratic discriminants and succeeded in providing formulations for positive definite and indefinite kernels. For biomedical signal analysis, a new clinical kernel function has been proposed in order to accurately represent similarities between patients and the kernel testing has been successfully carried out for the dataset of specific patients. This work indicates that adaptive kernels are needed to efficiently represent and classify clinical data and these in turn will aid in clinical decision support. A detailed review of some of the recent contributions to TF approaches can be obtained from the works of Shafi et al [33] and Sejdic et al [34].

Considering the work until now, it is interesting to note that, most of the existing works were targeted towards obtaining good representation for given signals. Very few works [31,35] have tried to exploit to information in the TF plane for obtaining good characterization for NS signals.

2.4   Quadratic TFD Selection

The required properties for an ideal quadratic TFD are: Non-negativity, realness, energy conservation, TF marginals, TF invariance, TF support, provision for instantaneous computations (time delay and instantaneous frequency), and reduced interference. However, in order to suitably represent a signal in the TF plane and depending on the application, a TFD need only satisfy a subset of the above mentioned properties (usually referred to as the "desirable" properties).

2.4.1   Desirable Properties

For a TFD to facilitate efficient characterization during NS signal processing, the following properties need to be satisfied:

1. **Realness:**

   Energy is a real quantity, so $A(\tau, \theta) = A(-\tau, -\theta)$, where $A(\tau, \theta)$ represents the ambiguity function defined by time-lag (τ) and frequency-lag (θ) components respectively.

2. **TF Invariance:**

   be invariant to time and frequency shifts. Also, changes in amplitude should not affect the performance of the TFD during feature extraction. In other words, the corresponding TFR should offer robust performance to time- and frequency- shifts. For any given

signal $x(t)$ and its shifted version $s(t)$,

$$x(t) = s(t - t_0) \; ; \qquad P_x(t, f) = P_s(t - t_0, f), \tag{2.8}$$

$$x(t) = s(t)\mathrm{e}^{-jf_0t} \; ; \qquad P_x(t, f) = P_s(t, f - f_0). \tag{2.9}$$

3. **TF resolution:**

   provide high TF resolution and allow computation of instantaneous spectral measures. A good TFD should offer maximum possible resolution while accommodating Heisenberg's uncertainty principle [36].

4. **TF Marginals:**

   have correct time- and frequency- marginals. For a given signal $x(t)$, its Fourier transform $X(f)$, and the corresponding TFD $P(t, f)$,

$$\int_{-\infty}^{\infty} P(t, f)df = |x(t)|^2, \tag{2.10}$$

$$\int_{-\infty}^{\infty} P(t, f)dt = |X(f)|^2. \tag{2.11}$$

   Correct marginals ensures high TF localization and offers good resolution for the obtained signal estimate.

5. **TF support:**

   provide good time (and frequency) support: the distribution should be zero before the start of the signal and after the end for finite duration signals. That is,

$$P(t, f) = 0 \text{ for all } t \text{ outside } (t_{start}, t_{end}).$$

In addition, some of the nice-to-have features are high clarity of auto-terms with no cross-terms, and lower computational complexity. Our investigation on the usability of these terms

will be discussed in detail in Chapter 3.

2.4.2 Illustration of Existing TFDs

In this section, the performance of certain quadratic TFDs were evaluated for assessing their potential for NS signal analysis. The following synthetic signal, $x(t)$, that is constructed as the sum of an increasing chirp and a decreasing chirp is used for obtaining an analytical comparison between TFDs.

$$x(t) = \sin\left(2\pi\left(f_0 t + k_0 * t^2/2\right)\right) + \sin\left(2\pi\left(f_1 t + k_1 * t^2/2\right)\right), \tag{2.12}$$

where, f_0, f_1 are the normalized (with respect to sampling frequency) characteristic frequencies of the time-varying signal and k_0 and k_1 are scalars. The signal in Eqn. 2.12 is obtained from two chirps: 1st whose normalized frequency decreases from 0.3 to 0 and an increasing chirp with 0 and 0.5 as the initial and final normalized frequencies respectively.

The illustration of the representations obtained from different TFDs, for a time-varying multicomponent input, is shown in Figure 2.3. It can be seen that, in comparison to the linear (Spectrogram and Gabor) distributions, the quadratic (WVD, CWD and RID) TFDs have better control over localization of time and frequency components. Further, WVD provides the best TF resolution; however, there are some additional energy components (that do not belong to the signal) located about the desired representation. These artifacts correspond to the terms generated during frequency component interactions between the two chirps. Although the CWD and RID's TFRs succeed in reducing the cross-term concentration, this comes at the expense of reduced TF resolution resulting from smearing. The RID provides the minimal cross-term concentration along with reasonable TF resolution (which is still weaker than WVD). Analysis reveals that the quadratic classes of distributions are suitable for NS signal analysis, both for representation and characterization purposes.

Choice of TFD for Characterization:

The spectral content of most real-life (natural and man-made) signals changes so rapidly that choice of a suitable short-time window that offers best preservation of the inherent TF connection in STFT energy representation (spectrogram) is problematic. Hence, for a given window, STFT has a fixed resolution i.e. either good frequency resolution or good time resolution in accordance with Heisenberg's uncertainty principle.

Although, scalogram (using wavelets) was introduced to provide adaptive resolution, it suffers from poor time resolution at low frequencies and poor frequency resolution at high frequencies. Thus, the resolution of the WT does not depend on the wavelet, but on the region of the time-scale plane at which the WT is computed [37]. Also, scalogram is translation variant, making it less preferable for quantification applications relating to existence of transients and NS input behavior.

The Gabor distribution offers the lowest TF resolution and localization, since it is challenging to model time-varying inputs using only Gaussian distributions. This makes the linear TFDs less useful for NS signal characterization. The smoothing distributions (CWD and RID) provide reasonable trade-off between auto-term concentration and reduction of cross-terms. However, their kernel functions manage only to suppress the interference terms which is away from the origin i.e. cross-terms located on the τ and θ axes cannot be removed by using kernels.

Although the quadratic TFDs suffer from cross-terms, owing to the bilinear transformation, they characterize the signal more efficiently than their linear counterparts and also, allow different realizations serving a variety of applications. Most importantly, the joint TF analysis techniques are non-parametric and hence, do not assume any signal model apriori, making them ideal candidates for time-varying signal characterization.

Table 2.1 shows the comparison of some of the available TFDs [38] in terms of the desirable properties for quantification applications. It can be noted that WVD satisfies most of the
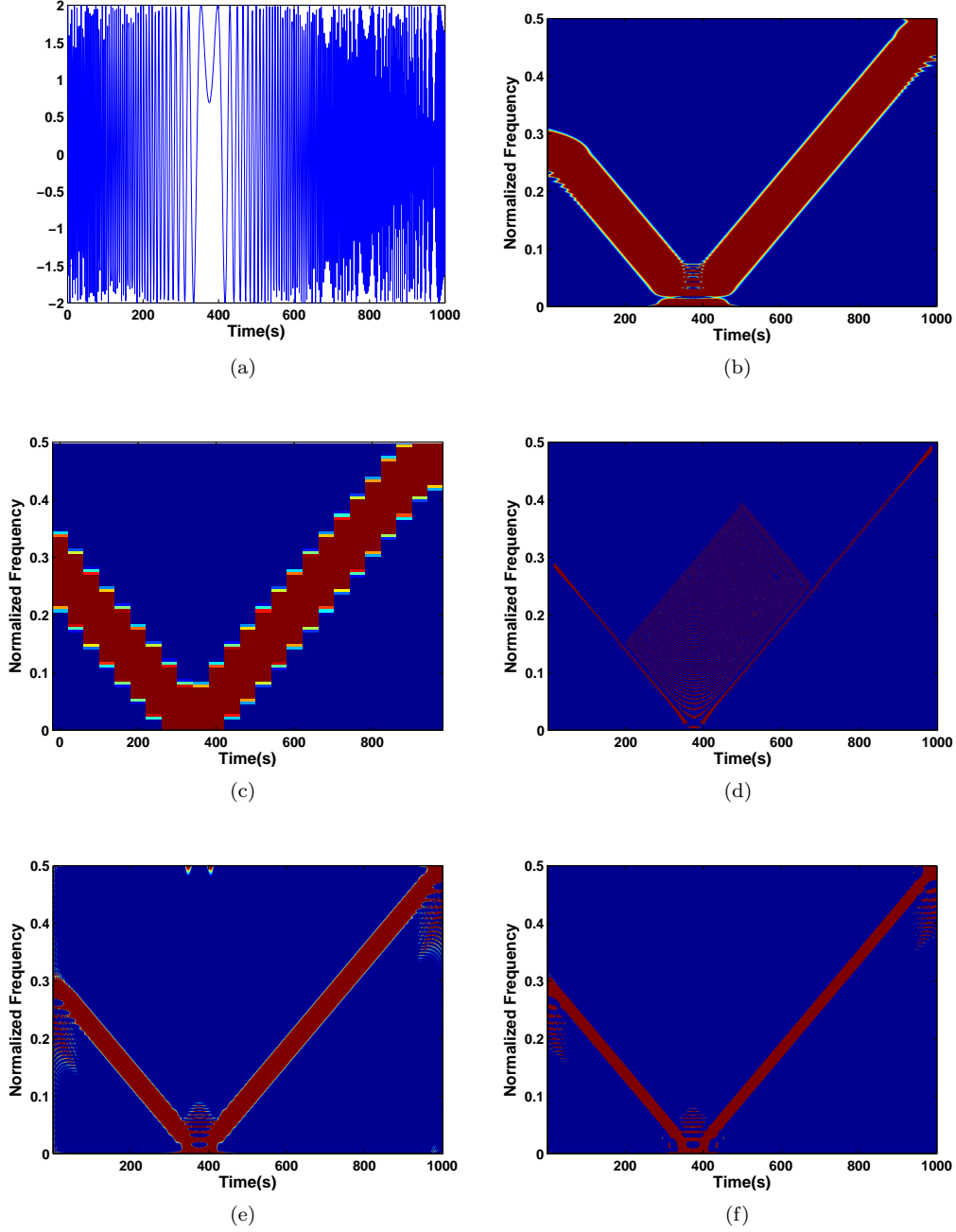
Figure 2.3: Illustration of different TFRs for a chirp input. (a) Input signal in time-domain. (b) Spectrogram with fast Fourier transform (FFT) size of 1024 points and using Kaiser-Bessel window function. (c) Gabor Representation using Gaussian window with range limiting parameter of 0.005. (d) The WVD with number of frequency bins equalling the signal length. (e) The CWD using an exponential kernel with the kernel width, $\alpha = 1$. (f) RID with a 1st order Bessel kernel and using a Hamming window for TF smoothing.

properties of an ideal TFD except for negative energy and the presence of cross-terms. Although the other available smoothing distributions (including RIDs) provide good cross-term suppression, they come at a cost of reduced resolution and poor localization of signal terms. Hence, there exists a trade-off in TFD performance owing to the tradeoff generated due to the amount of cross-terms. However, these (cross-) terms must be present for certain critical properties to be satisfied: correct marginals, instantaneous frequency (since it is calculated from the instantaneous autocorrelation function) and TF localization.

Based on the investigation on properties and WVD is the only distribution that offers good resolution and good energy compaction. For a WVD, the kernel function is an all-pass filter, $\phi(\theta, \tau) = 1$, [39]. Owing to the choice of the kernel, a given signal can be represented in two inter-related domains: either in TF domain or in the intermediate ambiguity domain formed by the TFD characteristic function. The choice of the intermediate domain (AD) is because of the following reasons:

(i) WVD does not satisfy non-negativity (of energy) property. Since energy cannot be negative (and WVD suffers from negative values for energy coefficients), the transformation is equal to the signal's ambiguity function (AF) and

(ii) In a multicomponent signal, in addition to signal terms, the cross-terms are generated due to the frequency interactions. Analysis of the cross-terms can thereby facilitate understanding of the underlying interactions and can in turn aid in characterization of time-varying signals. Also, the AD is an information rich domain mainly because of the unique positioning of signal and cross-terms in the AD-plane.

Considering the above mentioned benefits of the WVD and the AD, a novel scheme that exploits the AD map for analysis of the generated set of signal and cross-terms is introduced for obtaining robust characterization approaches for time-varying signals.

Table 2.1: Comparison of the properties of kernel-based TFDs with reference to signal characterization. RE-realness of energy; TM and FM - time and frequency marginals; IF - instantaneous frequency; TD - time delay; TS and FS - time and frequency support

Distribution	Property						
	RE	TM	FM	IF	TD	TS	FS
WVD	Y	Y	Y	Y	Y	Y	Y
CWD	Y	Y	Y	Y	Y	-	-
Spectrogram	Y	-	-	-	-	-	-
Rihaczek (Product kernel)	-	Y	Y	-	-	Y	Y
RID	Y	Y	Y	Y	Y	-	-

2.5 Chapter Summary

A concise overview of some of the well known linear and quadratic TFDs were presented in this chapter. The points of discussion were limited to the properties of desirability from a signal characterization perspective. Analytical evaluation of the available TFRs indicate that the available studies were limited to obtaining good representations and very few works [31, 35] target development of robust characterization strategies. Comparison of the TFD properties reveal that WVD is the only distribution that provides good TF localization and maximal TF resolution, although with the inclusion of cross-terms.

In a given time-varying multicomponent signal, all the generated terms can carry information in regards to the underlying behavior. Hence, a scheme that can quantify these terms can prove beneficial for extraction of characteristic features, in order to obtain automated analysis. In this context, the AD characteristics can be useful mainly because of the unique positioning of signal and cross-terms. A detailed investigation on the characteristics and usability of this domain, along with a novel time-varying characterization scheme is presented

in the next chapter.

Chapter 3

Methodology I: Signal Analysis in Ambiguity Domain

Owing to the time-varying nature of real-life signals, joint-variable techniques have gained increased importance among researchers. The evolution of Cohen's class of quadratic TFDs set new standards in NS signal processing. This is because, the time frequency (TF) plane preserves both temporal and frequency information in a given signal. The generalized expression for Cohen's class of TFD, $C(t, f)$, is given as [14],

$$C(t, f) = \frac{1}{4\pi^2} \iint M(\theta, t) e^{-tj\theta} e^{-\tau jf} d\theta d\tau, \tag{3.1}$$

where θ and τ represents the frequency and time equivalents in the ambiguity domain (AD). Here, $M(\theta, \tau)$, is the characteristic function of a TFD.

$$M(\theta, \tau) = \phi(\theta, \tau) \, . \, A(\theta, \tau). \tag{3.2}$$

For a given kernel ϕ, the characteristic function is defined as the dot product between the kernel function and the ambiguity function (AF), A, as given by 3.2. The kernel function
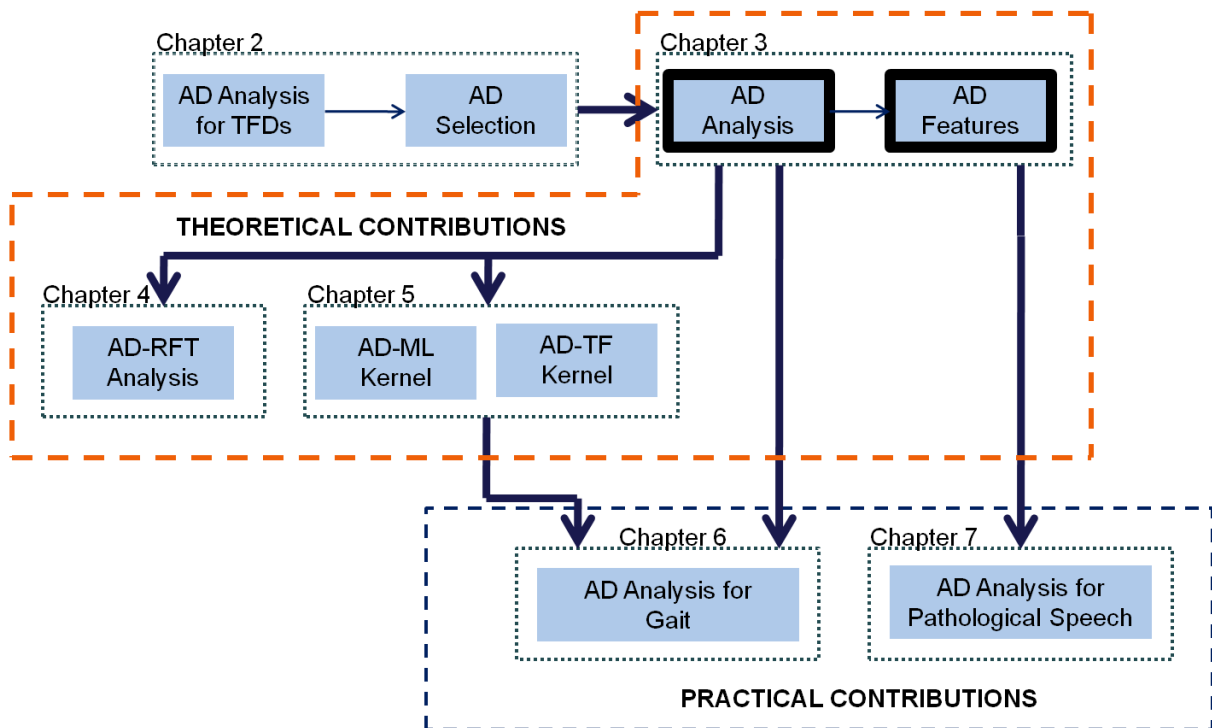
Figure 3.1: Chapter 3 - Signal Analysis in Ambiguity Domain

provides the 2-D filtering of the instantaneous autocorrelation function (A). A modification in the characteristic function (by using different kernels) facilitates design of different TFDs. Applications related to NS signal analysis have invariably considered the TF domain. This comes verily from the fact that TFDs provide reasonable performance accuracy, and allow instantaneous feature extraction for a variety of time-varying signal analysis. Investigation of such schemes indicates that the AD is an intermittent TF map. And the performance of a given TFD is usually modified by controlling the weighting factor (kernel) on the corresponding AF. This factor thereby controls the tradeoff between TF resolution and localization. Also, the AD representation, obtained by mapping the signal using the AF, provides the correlation map of a signal with its time-delayed version.

Research on AD indicates that, if not for radar signal applications [40] and optical interferometric image analysis [41,42], not much work is available that explains the usability of this domain for visualization and characterization of NS signal data. Garcia et al [43] proposed the use of an AD-based kernel for EEG signals and reported a minimization in the overall classification error. Details regarding the AD, along with its characteristics and significance are explained in this chapter as shown in Figure 3.1.

3.1 Ambiguity Domain

For a given TFD, the AD characteristics are controlled by the Fourier transformation of the 2-D kernel function. Also, literature [44] shows that the the second-order conditional moment in frequency of the TFD, defined by the corresponding AF, is proportional to the Teager Energy (TE) operator of a signal. From Eqn. 3.2 it can be verified that the characteristic function of a TF distribution, is proportional to the AF. This indicates the suitability of AD for analyzing the time-varying spectrum for NS signals. In addition, AD representation can be obtained from a signals' TFD by reverse-mapping the coefficients from the TF plane

i.e. the process is reversible and the TF representation can be easily obtained from AD and vice-versa.

3.1.1 Characteristics of AD

Figure 3.2 shows the relation between the time-domain signal, AF, Fourier transform, autocorrelation (spectral and instantaneous) and WVD. The rectangular blocks correspond to the domains that were previously used mostly for blind source separation and certain other radar applications. The shaded rectangular boxes list the domains that work with specific signal types and the oval-shaped box correspond to the widely exploited WVD (TF domain). Each domain possesses unique features in terms of their representability for certain types of signals. The transformation of the signal, $x(t)$, to autocorrelation representation, $R_f(t, \tau)$, is an irreversible process. It can be observed that since WVD uses an all-pass kernel, it does not carry any phase information and hence, the signals' phase information cannot be recovered. Hence, the signal estimate obtained by reverse mapping the TFD will not retain the phase information.

The choice of an all-pass kernel, results in generation of cross-terms and these terms introduce the negativity in magnitude for WVD.

By examining the relation between the TF space and the AD, it appears that some of the key features can be extended for both the domains. That is to say, the distinct advantages associated with TF analysis, such as localization of spectral components for extraction of instantaneous frequency and group delay features for a stochastic signal, are applicable for AD based analysis.

Properties of Ambiguity Function

The AF along with the kernel function defines the characteristic function of any TFD and hence the properties of the distribution holds valid for the AF as well. Two of the desired

Figure 3.2: Relationships between the time-domain signal, AF, Fourier transform, autocorrelation (a - spectral and instantaneous) and WVD

properties, that are essential for NS signal processing are satisfied by AF and they are defined as follows [14, 45]:

- **Time and Frequency shift invariance :** The magnitude of the AF remains unchanged irrespective of the signal shift in the time- and frequency- plane. However, there is a noticeable change in the phase factor due to modulation. In general, let $x(t)$ and $y(t)$ denote the input and output signals through a system.

$$y(t) = x(t - t_1) \, e^{j \, 2\pi \, v_0 \, t_0}. \tag{3.3}$$

Then, the corresponding AF matrix is obtained as,

$$A_y(\theta, \tau) = A_x(\theta, \tau) \, e^{j \, 2\pi \, (v_0 \tau - \tau_0 \theta)}. \tag{3.4}$$

- **Time and frequency Marginals:** The temporal and spectral auto-correlations are given by the points of intersection along the τ and θ axis respectively and the governing equations are given below:

Time Marginal

$$r_x(\tau) = A_x(0, \tau). \tag{3.5}$$

Frequency Marginal

$$R_x(\theta) = A_x(\theta, 0). \tag{3.6}$$

Maximum energy (obtained at the origin)

The energy of the input signal is given by the value of AF at the origin.

$$E_x = A_x(0, 0). \tag{3.7}$$

The AD can play a key role in analyzing the time-varying spectrum of a NS random process in both continuous and discrete time. For a given TF distribution, the kernel function is defined in the AD and is visualized as a 2-D filter. Based on the available literature [2, 14, 39, 40, 44] and our analysis about the AD, the following significant characteristics can be deduced.

1. The AF is a 2-D function of frequency-lag and time-lag components obtained from the signal. The intermittent map derived from any TFD by mapping the signal in the higher dimension is termed as the AD-plot.

2. As can be verified from Eq. 3.2, since AF defines the characteristic function of its governing TFD, there exist a reversible relation between the AD-plot and the corresponding TFR.

3. The signal mapping on the AD follows a specific pattern whereby the signal terms are grouped near the origin and the other terms resulting from multi-frequency interaction is spread out towards the edges of the plane. The generalized structure of the signal distribution is usually a bell or Gaussian shaped pattern.

4. The auto-terms fall on the time-lag axis for the case of narrow band signals and for wide band signals, they are spread over the entire domain. However, the specific positioning of auto- and cross-terms still hold good. This is because, the auto-terms form the diagonal terms in the signals AF matrix and the cross-terms form the off-diagonal entries.

Auto-Terms and Cross-Terms

For any bilinearly transformed multicomponent signal, the resulting TF distribution is comprised of two terms:

- *Auto-terms*: The signal terms corresponding to each of the TF pairs are commonly referred to as the auto-terms. For the case of narrowband signals, their corresponding auto-terms fall on the time-lag axis and for wide band signals, they spread over the entire AD.

- *Cross-terms*: Cross-terms correspond to the artifacts that falsely indicate the existence of signal components. These terms form the off-diagonal entries in AF-transformed signal matrix, so that auto-terms are positioned around the origin and the cross-terms are spread outside the signal terms. A scant number of researchers [40, 43] reported the use of this property for applications such as source separation and phase estimation in optic interferometric images [41, 42]. If not for radar applications, not much reference is available in regards to characteristics and applications of this domain.

  Although for obtaining better interpretability, the cross-terms are suppressed, these terms carry significant information regarding the relation between the different components. For example, in radar and sonar applications, the cross-terms represented in AD, aids in tracking of the moving target.

  All existing works target removal of these terms and in this research, an analysis scheme that exploits the cross-term distribution for NS signal characterization is proposed.

3.2   Proposed Approach

Research in the field of NS signal analysis involves two main objectives:

1. suitably defining the time-varying spectrum in terms of its distribution and

2. estimate of the relevant instantaneous parameters.

In addition to the above mentioned objectives, most existing TF techniques were developed for cross-term removal to enhance visualization in the TF-plane. This work is targeted
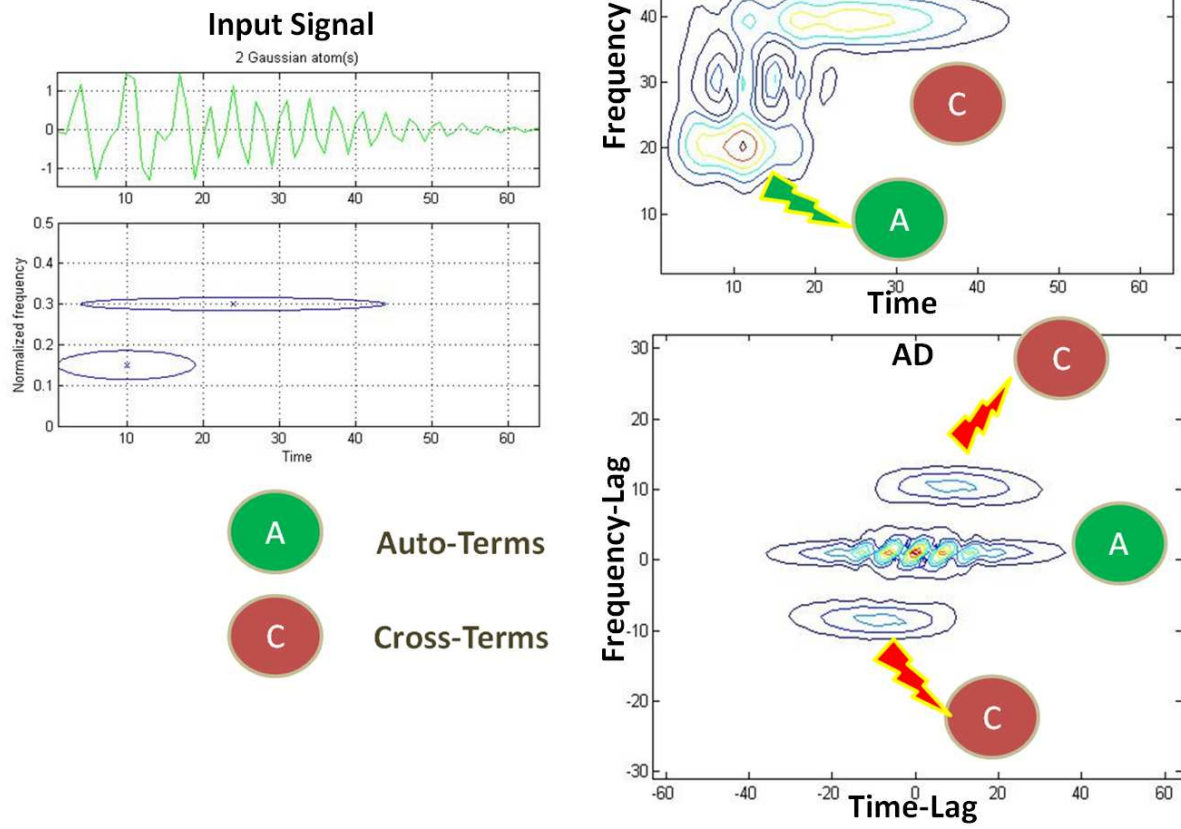
Figure 3.3: Left: Original Gaussian Atom, Right: TF plot (top) and AD-map (bottom) of the Gaussian atom

towards obtaining an efficient signal characterization scheme for NS signals.

Here, the AD space, a domain similar to the TF, is exploited to facilitate robust feature extraction and signal characterization. The current research work aims at suitably employing AD for design of discriminative analysis scheme for a variety of synthetic datasets and biomedical signals. As can be verified from Eqn. 3.2, the characteristics of the AD are defined by the AF. The AF is the time-varying correlation function, and in simple terms, it represents the energy as a function of time delay (τ) and the Doppler frequency (θ). For a given signal, $x(t)$, the AF is the integral,

$$A(\theta, \tau) = \frac{1}{E} \int_{-\infty}^{\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-j2\pi\theta t} dt, \quad (3.8)$$

where E is the total energy of $x(t)$ and is bounded for the input signal (usually normalized to unity).

The proposed AD-based characterization scheme aims at achieving the following significant objectives:

- Usability for study of auto- and cross-term energy for characterization.

- Analysis of the different signal components from the AD-map

- Detailed investigation on the usability of AD for NS signal processing, and

- Suitability for discrimination of biomedical signals.

A simple block diagram of an AD-based classification scheme is given in Figure 3.4. The input signal is normalized and then mapped on the AD-plane, using the AF (Equation (3.8), in order to obtain a transformed multivariate signal. Such a transformation aids in efficient capture of the signal characteristics in the higher order domain. Since AD is a symmetric correlative domain, in addition to signal terms, pairs of conjugate cross-terms corresponding to each signal interaction results. Hence, for a given N X 1 length input signal, the $N/2$ X

N set of coefficients from the AD-plane contains the same information as the entire domain. Also, since AF separates out the cross-terms as off-diagonal entries, the AD-mapped signal is useful for de-noising and source separation applications, where the cross-terms are filtered out using a low pass filter (suitable kernel function). The AF-half matrix is then used to define the feature space for classification application.

3.2.1 AD-based signal representation

In order to understand the characteristics of AD , certain stationary and NS signal examples are represented in the AD-plane. This assessment uses three signal types: periodic input (sinusoid), multicomponent time-varying input (Gaussian atoms) and a random signal (Gaussian noise).

(1) *Sinusoidal Input*:

The input and the corresponding AD plot of a sine wave are shown in Figure 3.5. As expected, the auto-terms appear in the time-lag axis, which is evident from both the surface and the contour plots. Internally, the auto-correlation of the sine wave with the delayed version of the same signal is computed during the transformation in the AD. In this case, only one cycle of the signal is considered and hence, there exists a signal spread on the time-lag axis. Further analysis revealed that, as the number of cycles increase, the spread decreases and a more concentrated gaussian lobe results for periodic inputs. Since a sinusoid is a single component signal, no frequency interaction terms (i.e cross-terms) are generated and this can further be verified from the obtained contour and surface plots.

(2)*Gabor Input*:

In order to assess the performance of the proposed technique for time-varying inputs, a Gaussian or Gabor atom input is considered next. This signal is unique in the sense that it has the greatest concentration in both the t and the f domains and the analysis of such a signal is expected to provide a wide range of perspectives about the signal representation in

Intermittent TF Representation
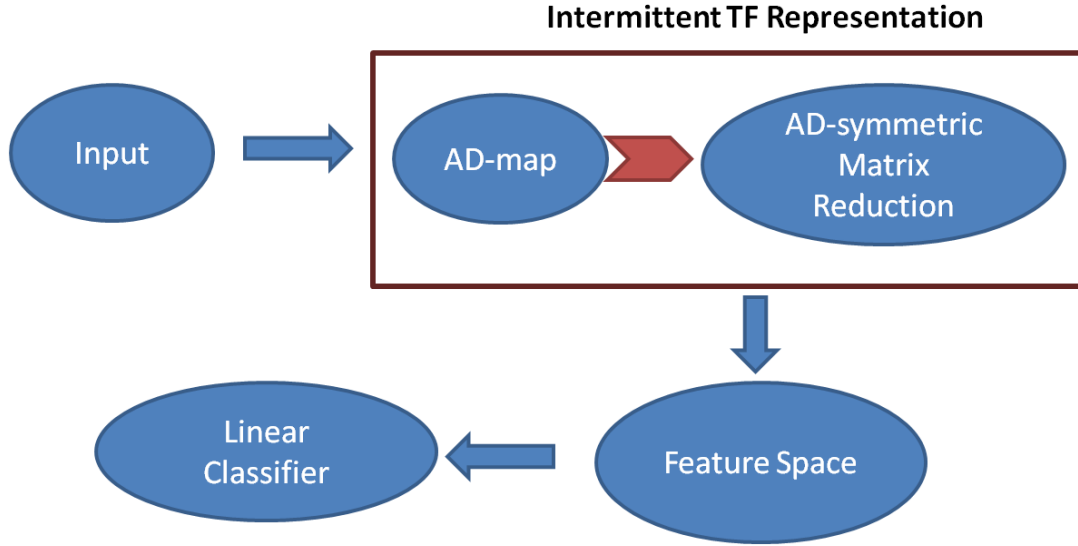
Figure 3.4: Block diagram of a AD-based classification scheme

the AD. A 3-component Gabor atom is used in this analysis and is as shown in Figure 3.6. From Figure 3.7, it can be verified that as expected the auto-terms, like for the case of sine wave, are concentrated around the origin. Since the Gaussian input (Figure 3.7) is correlated to some extent, the auto-terms are represented by the bell-shaped curve of maximum energy centered about the origin and six prominent conjugate-symmetric cross-terms positioned outside the region of auto-terms.

In addition, there exist three complementary cross-term pairs, corresponding to the interaction between the dominant frequency components with the remaining components. It is also interesting to note that owing to the symmetric behavior of AD, the cross terms appear in conjugate pairs and appear away from the auto-terms. From the simulation results, obtained from the sine wave and Gaussian atom, it is evident that for applications involving denoising, a 2-D kernel (to function as a low pass filter) can be used to separate out the signal terms from the interfering signals.

In general, for a m component signal, m signal components and $m-1$ cross-terms result. For
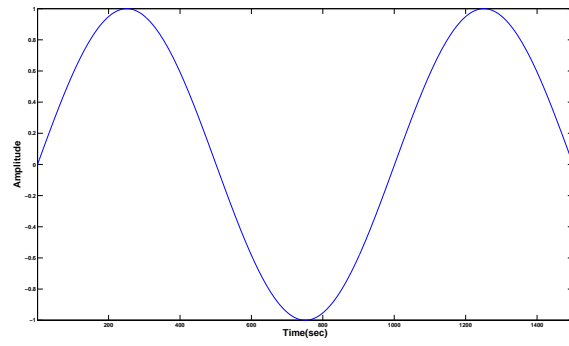
(a) Sinusoidal Input
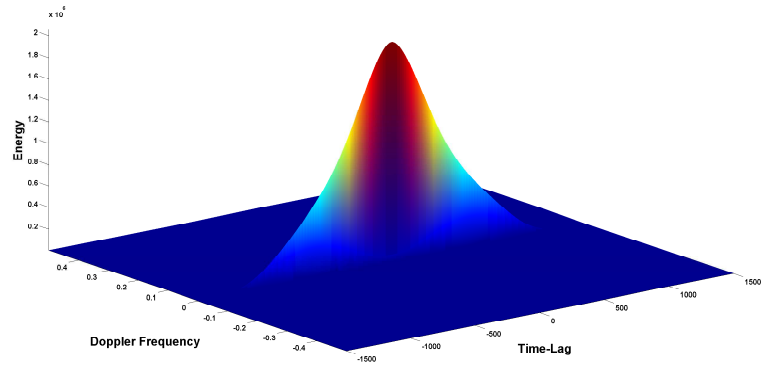
(b) Surface AD plot of a sine wave input
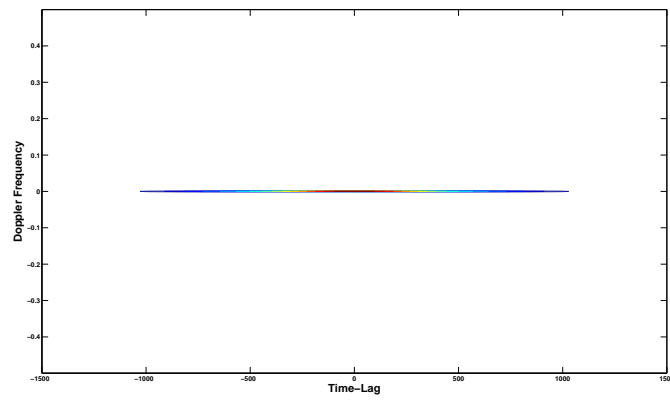
(c) Contour AD plot of a sine wave input

Figure 3.5: (a) A Sinusoidal Input and its corresponding (b) surface and (c) contour plots in the AD
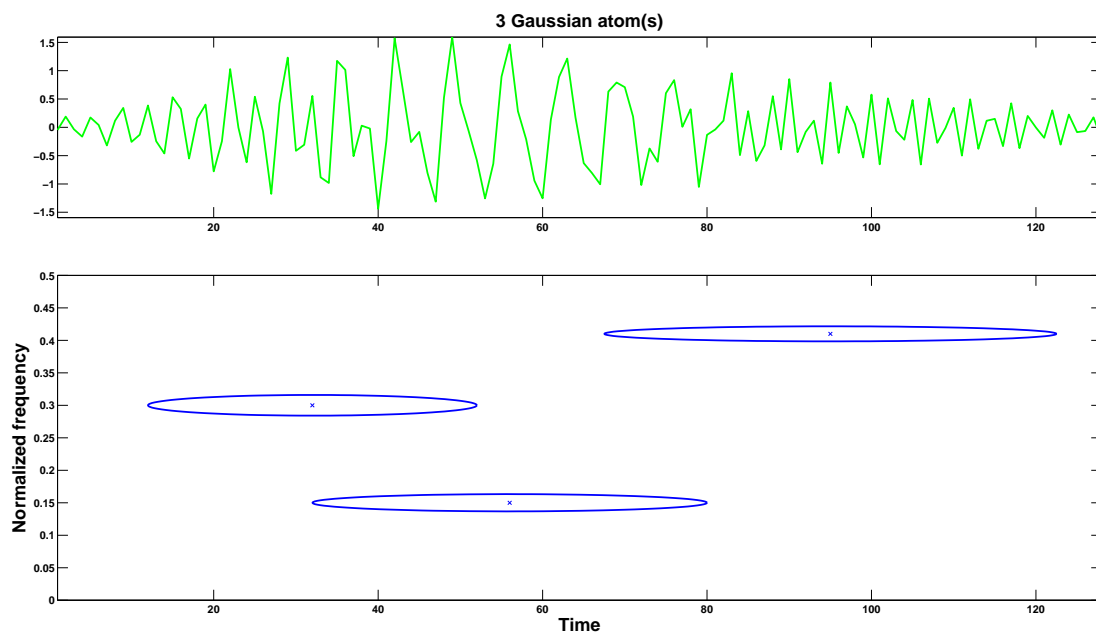
Figure 3.6: Top: Gaussian input and Bottom: the corresponding TF representation
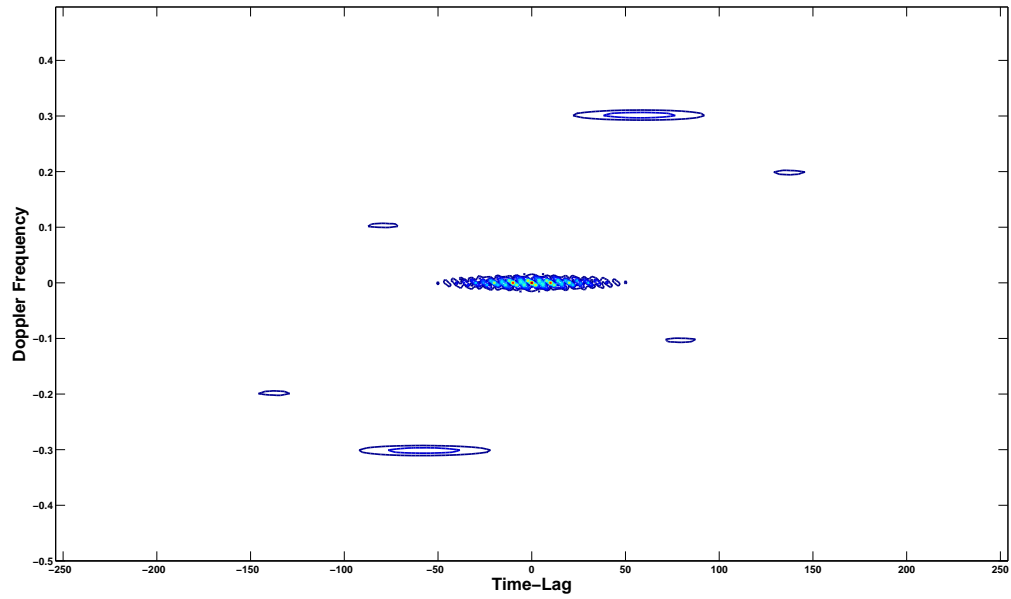
(a) Contour AD plot of a Gaussian atom
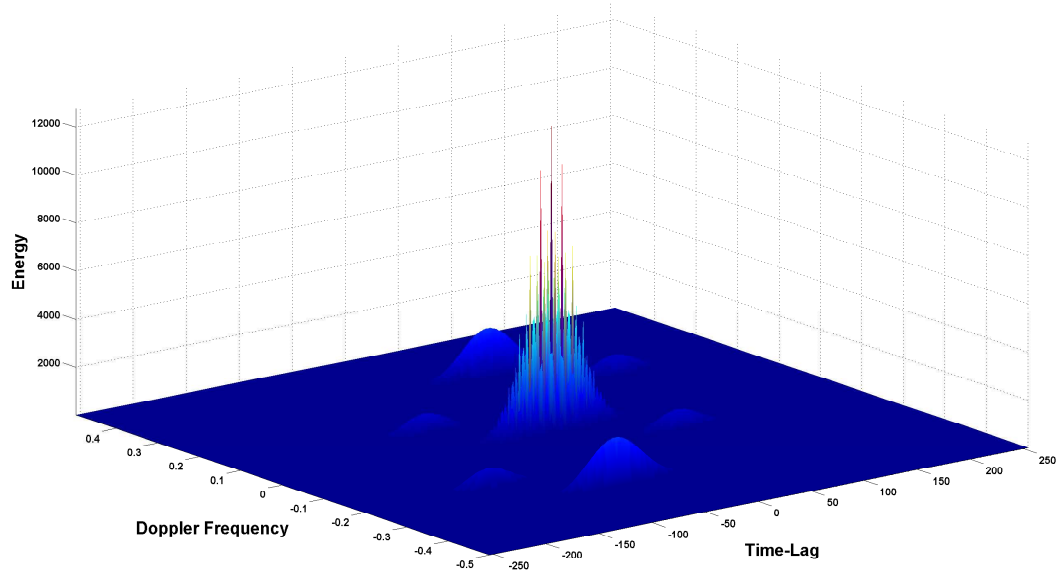
(b) Surface AD plot of a Gaussian atom

Figure 3.7: AD representation of a 3-component Gaussian signal. (a) Contour AD plot and (b) Surface AD plot showing the auto-terms and the cross-terms.

extreme scenarios, when the signal-frequency components are located close to each other, there is a good probability that some (or all) of the generated cross-terms overlap with the signal region resulting in distorted output.

(3) *Random Noise*:

The third signal under analysis, is a noise input. For any given signal, the obtained signal-AF matrix have maximum values along the diagonal and is symmetric about the origin which can be verified from the results obtained from Gaussian atom representation. The structure (or auto-term energy) spread depends on the temporal coherence in the signal. The more disoriented a signal is, the less will be its spread and this can be verified from the AD-map for a Gaussian noise as shown in Figure 3.8. Because of the presence of high randomness, the AD-plot has a single peak at the origin (i.e. at zero shifts, for all other integer shifts the autocorrelation function output is close to zero). Also, the location of cross-terms with respect to origin is determined by the closeness in frequency between the generated cross-term and the signal terms.

3.2.2 Feature Space

The AD-map of the transformed signal provides the energy distribution among the auto- and cross-term components. To facilitate characterization using the proposed scheme, a novel energy-based feature extraction technique is used in conjunction with the AD-representation. The feature space is defined by considering the energy distribution map across the signal- and the cross-terms generated due to interaction between different frequency components. The choice of energy-based feature extraction from the AD-space, is triggered by the distinct representation on the AD i.e. on account of the separation that an AD-map offers for the terms generated using a bilinear transform. It has been previously verified that on the AD-plane, the auto-terms of a time-varying multicomponent signal are located around the origin and the cross-terms are placed away from the origin at a distance function defined by
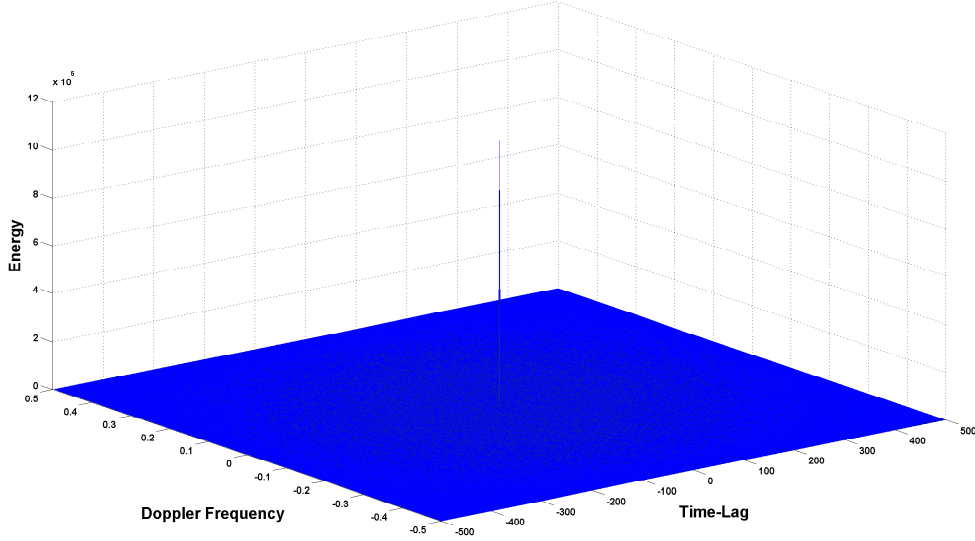
Figure 3.8: AD-plot of Gaussian noise signal

the separation between different components in time and frequency axes. This important property is a key idea involved in the design of the proposed information extraction scheme using AD.

Prior to defining the feature plane, the AD-map is divided into 'n' equal bands as shown in Figure 3.9. For a given signal length $N \leq 2^m$, the number of bands is given by,

$$n = \text{floor}(m/2).$$

In order to efficiently quantify the energy distribution across these energy bands, a novel Relative Distribution Ratio (RDR) for each segment (i = segment number) is defined and is given as,

$$RDR_i(n) = \frac{1}{nT_E} \left\{ \sum_{k=1}^{n} [A_k(\theta, \tau) - A_i(\theta, \tau)] \right\}. \tag{3.9}$$

Here, $A_n(\theta, \tau)$ represents the total energy distribution in the nth band i.e. total energy for

50
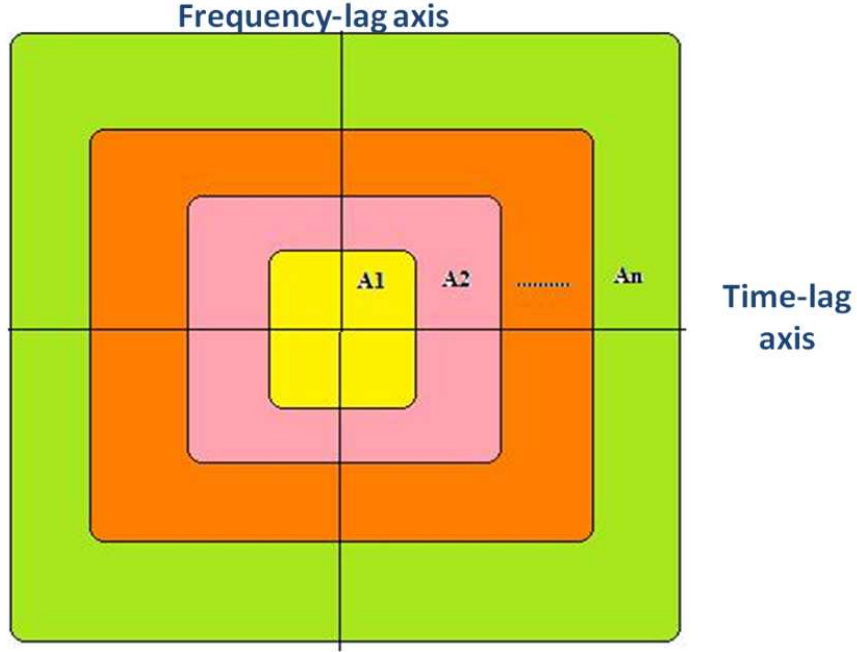
Figure 3.9: Partition of AD-plot into relative energy bands

all values of θ and τ in the nth energy band. This process ensures the computation of energy spread from low-frequency to high-frequency region. The spread can then be computed to study the distribution in detail for different classes of NS signals. In Eqn 3.9, T_E represents the energy in the innermost bands around the origin. Based on investigation of the AD-map of 20 multicomponent synthetic signals, it is interesting to note that over 75% of the energy is distributed within the inner three bands. Since auto-terms form the diagonal entries and are positioned near the origin, the value of T_E quantifies for most of the auto-term energy. Hence, the inner bands will have maximum auto-term energy in comparison to outer bands. The value of T_E can be computed from the following equation:

$$T_E = \sum_{\theta_1} \sum_{\tau_1} |A_i(\theta_1, \tau_1)|^2 \,, \tag{3.10}$$

where θ_1 and τ_1 takes the following values,

$$\theta_1 = \left|\left\lfloor \frac{r}{2} \right\rfloor \pm k_1\right| \text{ and } \tau_1 = \left|\left\lfloor \frac{c}{2} \right\rfloor \pm k_2\right|, \tag{3.11}$$

where the scalars k_1, k_2 define the regions of over 75% energy distribution (about the origin), r x c gives the size of the AF matrix and $\lfloor . \rfloor$ denotes the floor function. From RDRs, certain feature vectors are defined by measuring relative statistics in different energy bands.

The RDRs correspond to the energy values and they facilitate relative quantification of auto- and cross-terms suitably. Since a given multicomponent signal is transformed bilinearly using the AF, a quadratic superposition principle is satisfied i.e. both 1st order and 2nd order computations are considered. And the presence of the 2nd order terms have a significant influence on the generated terms and hence, cross-terms appear. For instance, a N component signal when transformed bilinearly will result in N signal terms (grouped around the origin) and $N(N-1)$ cross-terms spread away from the origin in the AD-plane.

In general, the generated interfering terms are considered troublesome since, for multicomponent signals, these terms often hamper readability and are often misconstrued for presence of signal. On the other hand, the presence of these terms offers an improvement in the TF localization, when compared to linear transforms. Hence, quantifying their relative distribution can facilitate understanding of the mechanisms generating them, and is therefore considered while defining the feature space. The proposed RDRs can be used to provide the relative cross-term distribution directly, since the auto-terms are always distributed about the origin and the energy spread away from the origin correspond to the signal frequency interactions in the correlative AD-plane. This concept can be extended for characterization of different multicomponent signals, that differs in the number of dominant frequency components and also, can be used to understand the underlying dynamics of the signal generating source. The extracted set of novel feature vectors along with the class labels can be used in applica-

tions involving discrimination between different classes of data. Since, the choice of feature vectors cannot be made arbitrarily, certain application-specific feature vectors are introduced in the later chapters. Additional insight with reference to classification performances will be discussed in detail in Chapters 6 and 7.

3.2.3 Significance of Proposed AD-based approach

- **Ease for characterizing the auto- and cross-terms:**

  For the case of multi-component signals, AD shows great potential mainly due to the specific positioning of the auto- and the cross-terms. Such a representation can facilitate characterization of signals based on the relative energy distribution of the signal terms and their interactions.

- **Cross-term mapping:**

  Signal terms are generated during correlation of same signal components and cross-terms are generated when different components are correlated during transformation. Owing to the reversible relation between AD and TFD, certain auto- or cross-term components can be mapped back and forth for further analysis in domain of choice. Although exact separation of energy between auto- and cross-term spread is not feasible, the proposed RDRs can facilitate quantification of relative CT energy in TFD.

- **Faster Processing:**

  Since AD representation is obtained from an intermediate stage (before computing the TF distribution), FFT computation can be avoided with the number of computations reduced by $O(N \, log \, N)$) for a signal length of N. This is along with a reduction in the matrix dimension used in the analysis.

- **Potential for signal characterization:**

  Offers good potential for developing schemes for real-life signals (biomedical) NS signal

characterization and their analysis.

- **Facilitates cross-term reduction:**

  For certain applications involving analysis of time-varying spectrum without the presence of any interfering signal, AD offers feasible solution using simple low pass filter (kernel).

3.3   Chapter Summary

An AD-based characterization technique that retains the TF relation over the entire duration of the signal is investigated and exploited for obtaining a novel discriminative analysis scheme. A detailed investigation reveals the significant characteristics that this domain offers for signal characterization. Following this, a discriminative-yet-representative set of feature vectors, for AD-based characterization of the relative signal and cross-term energies, is proposed using the RDRs.

Further, a detailed investigation on the concept of cross-terms is presented. Although these terms are considered to plague the representation (i.e. results in distorted visualization), from a signal characterization perspective these terms carry information about the signal term interactions. These factors were considered and characterized using the proposed RDRs.

In the next chapter, the theory of AD-map is extended by proposing a novel transformation strategy that uses the number theory concepts for obtaining a distribution that offers inherent sparsity.

Chapter 4

Methodology II: AD Transformation using Ramanujan Sums

In the previous chapter, the concept of AD and its unique characteristics that makes this domain suitable for NS signal analysis were presented. In this chapter (Figure 4.1), an extension to the theory of TF analysis is provided in AD, using a novel class of signal transforms.

4.1 Motivation

Research in signal processing shows that a variety of transforms have been introduced to map the data from the original space into the feature space, in order to efficiently analyze a signal. These techniques differ in their basis functions, that is used for projecting the signal into a higher dimensional space. One of the widely used schemes for quasi-stationary and NS signals are the TF transforms, characterized by specific kernel functions.

Literature [14] in signal processing show that a plethora of approaches have been developed for complex data analysis. One of the widely used methods is the Discrete Fourier Transform
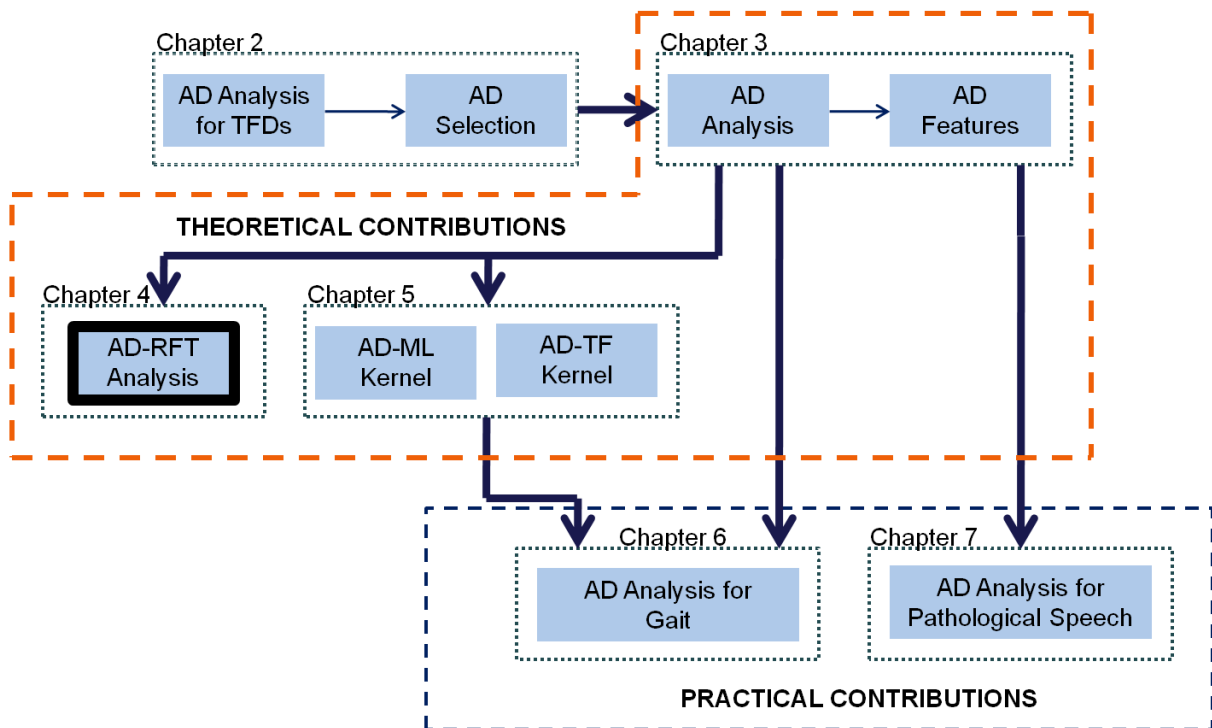
Figure 4.1: Chapter 4 - AD Transformation using Ramanujan Sums

(DFT). Throughout the scope of this work, the FFT, a faster algorithm to compute DFT (and its inverse), will be used for simulation purposes. Other complementary techniques such as wavelet analysis and empirical mode decomposition were developed to identify useful patterns from seemingly random sequences [46]. The main difference, between each of the available techniques, arises due to the basis function that is used for projecting the signal onto a different subspace.

Recently, the concept of Ramanujan Sums (RS) has been introduced in a signal processing perspective [47]. This is owing to the inherent orthogonal property of these sums to obtain convergent functions. Using number theory, these sums were later expanded to formulate the Ramanujan Fourier Transform (RFT) [47]. Literature [48] shows the usefulness of this class of transforms for achieving faster implementations in comparison to the DFT, since only the co-resonant frequency terms are used in the computation.

Cohen [49, 50] initially investigated the number-theoretic concepts to understand the usefulness of the so-called even functions for signal analysis. These integer valued weighting factors formulated by Cohen [49, 50] were later verified to be the RS originally derived in [51]. The usability of these functions for signal processing has been recently investigated. Planat et al [47] first used RFT as a tool for analyzing low-frequency noise in periodic, quasi-periodic and complex time series, as an alternative to FT. Following this, a RFT-based novel scheme is presented to facilitate analysis of the secondary structure content in amino acid sequences [52]. Their research showed that in the presence of Gaussian noise, RFT captures the signal periodicities better than the conventional FFT. Lagha et al [53] expanded on the theory of usability of this transform for estimation of Doppler spectral parameters for weather signals. The usefulness of this class of transform is also reported for analysis of biomedical signals, such as T-wave alternans, by Mainardi et al [54].

In this research work, an extension to the 1-D RFT into the 2-D space for signal analysis is proposed. The 2-D interpretation is then used to project the data from AD into the TF

space, in place of the standard FFT. Following this, the dimensionality reduction capabilities of the deduced transform, for certain signal classes, are investigated.

There are many different ways to describe a signal mathematically and the main motivation behind selection of a specific method comes from the choice of basis functions so as to suit the given signal and the corresponding application. For example, the choice of Fourier basis for stationary signals and the use of adaptive basis functions for time-varying signal analysis. One of the desired properties for a given set of basis functions is 'Finality of coefficients', which states that each of the coefficient can be determined independently without the necessity of availability of other coefficients [55]. Existence of orthogonality among basis functions, satisfies this property and hence, the potential of RS functions are investigated for TF analysis.

For analyzing certain quasi-stationary and NS signals, TF decompositions are largely preferred, owing to the flexibility these techniques offer for simultaneous computation of time-varying parameters. This work is driven by the fact that RFT decomposes a signal based on the co-prime resonances $(p, q) = 1$. That is, only those samples that are co-prime to the value of q will contribute to the basis function, and this indicates only selected few samples contribute to the sum. This is an interesting directive since noise-like terms will not contribute to the sum and the resulting spectrum will have less harmonics. One direct implication is that, only a reduced number of samples are considered for estimation. This characteristic offers certain degree of inherent sparsity from the transform, which when exploited suitably can aid in obtaining reduced dimensionality approximations for a given signal.

In addition, the RFT could operate on the entire signal length (for an infinite length series) [56]. In this research, the usability of RFT for obtaining a characterization map in the TF plane will be investigated and the suitability of this technique for other bilinear transforms, to analyze complex signals is evaluated.

4.1.1 Ramanujan-Fourier Expansion

Ramanujan sums [51] are real sums defined as the nth powers of the qth primitive roots of unity. In number theory, a number q is said to be the primitive root of unity if the powers of n include all the residue classes of modulo '1'. RS are defined as follows,

$$c_q(n) = \sum_{p=1;(p,q)=1}^{q} e^{\left(2i\pi \frac{p}{q}n\right)}, \tag{4.1}$$

where $(p,q) = 1$ indicates that the greatest common divisor (GCD) is unity i.e. p and q are co-primes. The RS is the basis function over which the signal $x(n)$ is projected. It is known, see for example, Apostol [57], that the RS are orthogonal in nature and hence offer good signal energy conservation for analysis, similar to the FFT.

These sums are used to define a different decomposition represented as,

$$x(n) = \sum_{q=1}^{N} x_q c_q(n). \tag{4.2}$$

In Eqn. (4.2), x_q's are referred to as the RFT coefficients. In this transform, the signal is projected on a set of basis functions defined by their corresponding RS which is obtained as multiples of a frequency $(1/N)$ defined by the signal length. By evaluating the function over $q \to \infty$, Eqn. (4.2) equals the Fourier series. From Eqn. (4.1) it is to be noted that only components that are co-primes contribute to the sum. Hence, by signal inspection followed by a reasonable choice of the upper bound for q, a reduced dimensionality decomposition can be obtained.

An alternative computation of RS can be obtained from the following relationship, using the Euler totient function $\phi(n)$ and Moebius function $\mu(n)$ [58].

$$c_q(n) = \mu\left(\frac{q}{(q,n)}\right) \frac{\phi(q)}{\phi\left(\frac{q}{(q,n)}\right)}. \tag{4.3}$$

In Eqn. (4.3), the Euler totient function, $\phi(q)$, is the multiplicative arithmetic function defined for positive integers q and is given by the number of positive co-prime integers less than or equal to q. The Mobius function $\mu(n)$ is also a multiplicative function and is zero for positive integers which are not square-free.

Also, from (4.3) it can be observed that [57],

$$
\begin{aligned}
\text{if } (q, n) &= 1, \ c_q(n) = \mu(q) \text{ and} \\
\text{when } (q, n) &= q, \ c_q(n) = \phi(q).
\end{aligned}
\tag{4.4}
$$

Using the relationship defined in (4.3), Carmicheal [59] formulated another representation for RFT which is given by the following equation.

$$
x_q = \frac{1}{\phi(q)} A_v(x(n) \, c_q(n)),
\tag{4.5}
$$

where, $A_v(g)$ is the mean value of the function $g(n) = x(n)c_q(n)$.

$$
A_v(g) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(n).
\tag{4.6}
$$

4.2 Properties of RFT

Following the introduction of RFT, a detailed investigation of the properties of RFT is performed. This section presents some of the deduced properties of RFT, including linearity, time-reversal, scaling, time-shift, energy conservation and the equivalence to wavelets.

1. Linearity

   For any two given signals $x_1(n)$ and $x_2(n)$ and their corresponding Ramanujan-Fourier

coefficients x_{q_1} and x_{q_2}, the discrete RFT is linear, if

$$\mathrm{R}(x_1(n)) = x_{q_1},$$

$$\mathrm{R}(x_2(n)) = x_{q_2},$$

$$\mathrm{R}(ax_1(n) + bx_2(n)) = ax_{q_1} + bx_{q_2}, \tag{4.7}$$

where a and b are complex constants and $\mathrm{R}(.)$ represents the RFT decomposition. During the scope of this work, $\mathrm{R}(.)$ and $\overset{R}{\rightarrow}$ are used invariably to represent the RFT decomposition.

2. Time-Reversal

The value of the Euler totient function used in Eqn. 4.4 is not time-reversible and has the same value for both positive and negative instants of 'q'.

$$\phi(q) = q^2 \prod_i \left(1 - \frac{1}{q_i^2}\right),$$

$$\text{and } \phi(-q) = q^2 \prod_i \left(1 - \frac{1}{q_i^2}\right).$$

This implies,

$$\phi(q) = \phi(-q). \tag{4.8}$$

Hence, the time-reversal property of RFT is similar to that of the conventional DFT. That is,

$$\mathrm{R}(x(n)) = x_q,$$

$$\mathrm{R}(x(-n)) = x_{-q}.$$

3. Scaling

$$\mathrm{R}(x(n)) = \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_n x(n) c_q(n) \right].$$

For a time-scaled signal,

$$\mathrm{R}(x(an)) = \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_n x(an) c_q(n) \right].$$

Let $m = an$; $n = \frac{m}{a}$. Then,

$$\mathrm{R}(x(m)) = \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_m x(m) c_q(m/a) \right]. \tag{4.9}$$

Similarly, for $x(n) \to x(n/a)$ and $m = n/a$:

$$\mathrm{R}(x(m)) = \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_m x(m) c_q(am) \right] \tag{4.10}$$

The RHS in the Eqns. (4.9) and (4.10) is simply the RFT of $x(n)$ evaluated at different m. However, not all samples are included in the computation and the value of the integer scaling factor a controls the same. And depending on the signal length n, the scaling process can lead to loss of input samples.

4. Time-Shift

$$\mathrm{R}(x(n)) = \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_n x(n) c_q(n) \right].$$

For a time-shifted input, i.e. $x(n) \to x(n - n_0)$,

$$R(x(m)) = \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_m x(m) c_q(m + n_0) \right],$$

$$\text{where } c_q(n) = \sum_{p=1;(p,q)=1}^{q} e^{2i\pi \frac{p}{q} n},$$

$$c_q(m + n_0) = \sum_{p=1;(p,q)=1}^{q} e^{2i\pi \frac{p}{q}(m+n_0)},$$

$$\text{and } c_q(m + n_0) = c_q(m) c_q(n_0).$$

Hence,

$$R(x(m)) = \frac{c_q(n_0)}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_m x(m) c_q(m) \right], \tag{4.11}$$

$$R(x(m)) = c_q(n_0) \left\{ \frac{1}{\phi(q)} \left[\lim_{N \to \infty} \frac{1}{N} \sum_m x(m) c_q(m) \right] \right\}. \tag{4.12}$$

A shift in time-domain corresponds to an equivalent shift in the compressed coordinates of the periodicity domain, however, the RFT coefficients are weighted by the factor $c_q(n_0)$.

5. Volume Invariance

For a given discrete signal and its corresponding RFT,

$$R(x(n)) = x_q.$$

Hildebrand et al [60] proved that, if $x \in R^2$ (modulo null-functions), then the even arithmetical functions (Ramanujan sums) satisfies the following Parseval's relation.

$$\|x(n)\|_2^2 = \sum_{q=1}^{\infty} |x_q|^2 \, \phi(q). \tag{4.13}$$

Here, the $|x_q|$ indicates that the Ramanujan coefficients are complex-valued, like Fourier coefficients. Eqn. (4.13) indicates that in order to satisfy volume invariance, the mean function (and in turn the Ramanujan sums) should exist (and be non-zero) for all considered values of q.

Although Ramanujan transform can efficiently capture the inherent periodicity of a given input (even under the presence of noise), Parseval's theorem fails if the maximum value of q is less than ∞, resulting in a signal approximation rather than a perfect decomposition. The ratio of the signal energy and the maximum energy of RFT-mapped signal coefficients converges to unity as $q \to \infty$.

6. Equivalence to Wavelets:

   For a square-free positive integer, the only values that the Moebius function (μ) can take are a series of ones (± 1). Hence, the special case (in 4.4) when the GCD equals '1' generates a series of positive and negative peaks. These signals can also be suitably used to define a dyadic Haar-like projection in the space-frequency space.

   The generated functions are not continuous (and hence, non-differentiable), making them usable to analyze signals with sudden transitions. The resulting filter function offer faster computations and a relatively simpler implementation structure.

4.3 Time-Frequency Analysis Using RFT

For analysis of certain quasi-stationary and NS signals, TF transformations provide reasonable characterization of the time-varying parameters and hence, is one of the widely preferred techniques. In the previous chapter, the usefulness of AD for NS analysis was investigated and in this section, scheme for AD analysis while using the RFT in place of the conventional FFT is introduced.

Figure 4.2 presents the proposed architecture for the alternative AD expansion. The TF
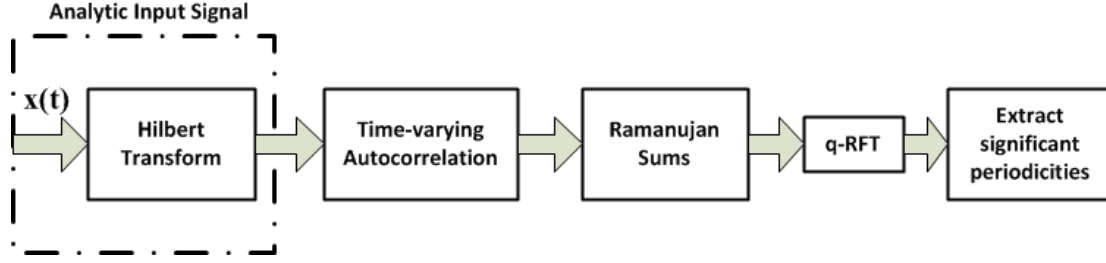
Figure 4.2: Block Diagram for TF-RFT Representation

representation is obtained by Fourier transformation of the characteristic function. The characteristic function is obtained as the dot product between the kernel function and the AF. In this work, the case of an all-pass kernel (Wigner-Ville distribution) [16], where the kernel $\phi(\theta, \tau) = 1$, is used. Here, the analytic input $x(t)$, obtained using the Hilbert transform module, is projected on the 2-D AD space, by using the AF defined as follows,

$$A(\theta, \tau) = \int x(t - \frac{\tau}{2}) \, x^*(t + \frac{\tau}{2}) \, e^{-j\theta t} dt. \tag{4.14}$$

The AF computes the time-varying autocorrelation function and owing to the bilinear structure of the transform, in addition to signal-terms, certain artifacts (or cross-terms) also results by interactions between signal components. A unique feature of the AD-plane is that, in this domain, the signal terms appear localized around the origin and the cross-terms are spread away from the origin [14]. In this work, a modification of the AF representation is computed by using the RS (on the time-varying autocorrelation values) as,

$$\hat{A}_q(\theta, \tau) = \frac{1}{N\phi(q)} \int_{t=1;(t,q)=1}^{q} \left\{ x\left(t - \frac{\tau}{2}\right) x^*\left(t + \frac{\tau}{2}\right) \right\} e^{j2\pi\frac{t}{q}} \, dt. \tag{4.15}$$

Such a computation results in an overall reduction in the number of coefficients in AD for two reasons: (i) primarily owing to the symmetricity of AD only half plane information would suffice and (ii) secondly since RFT operates only on co-prime resonances and is defined by the choice of q, proper selection of q can facilitate additional reduction in overall dimension.

65

These AD-mapped coefficients can then be transformed into a sparser TF representation, using the RFT computed from their characteristic RS.

4.4   2-D RFT

The 1-D RFT provided in (4.2) is expanded and a 2-D formulation is proposed. The 1-D RFT and the 2-D equivalent are defined in (4.16) and (4.17) respectively.

$$x_q = \frac{1}{\phi(q)} \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x(n) c_q(n), \tag{4.16}$$

$$x_{q_1,q_2} = \frac{1}{\phi(q_2)} \lim_{N_2 \to \infty} \frac{1}{N_2} \sum_{\tau=1}^{N_2} \{R_1\} c_{q_2}(\theta), \tag{4.17}$$

$$\text{here, } R_1 = \frac{1}{\phi(q_1)} \lim_{N_1 \to \infty} \frac{1}{N_1} \sum_{\theta=1}^{N_1} A_c(\theta, \tau) c_{q_1}(\tau),$$

where A_c represents the time-varying autocorrelation function, q_1, q_2 are the RFT lengths for each of the 1-D computations and N_1, N_2 are the lengths of row and column vectors respectively.

4.4.1   Algorithm: 2-D RFT of the AF matrix

1. In equation (4.17), the function R_1 in {..} brackets corresponds to the one-dimensional RFT of the n_1^{th} row and is computed from the standard equation (4.2).

2. Each row in the AF-matrix is then replaced by their corresponding RFT coefficients

computed in the previous step.

$$
\begin{bmatrix}
a_{11} & a_{12} & . & a_{1N_1} \\
a_{21} & a_{22} & . & a_{2N_1} \\
. & . & . & . \\
. & . & . & . \\
a_{N_21} & a_{N_22} & . & a_{N_2N_1}
\end{bmatrix}
\xrightarrow{R}
\begin{bmatrix}
a^r_{11} & a^r_{12} & . & a^r_{1q_1} \\
a^r_{21} & a^r_{22} & . & a^r_{2q_1} \\
. & . & . & . \\
. & . & . & . \\
a^r_{N_21} & a^r_{N_22} & . & a^r_{N_2q_1}
\end{bmatrix} .
$$

3. Following this, a column-wise repetition of Step. 2 is performed and in a similar fash-ion, the column vectors (for each of the n_2^{th} column) are replaced by their Ramanujan coefficient equivalents.

$$
\begin{bmatrix}
a^r_{11} & a^r_{12} & . & a^r_{1q_1} \\
a^r_{21} & a^r_{22} & . & a^r_{2q_1} \\
. & . & . & . \\
. & . & . & . \\
a^r_{N_21} & a^r_{N_22} & . & a^r_{N_2q_1}
\end{bmatrix}
\xrightarrow{R}
\begin{bmatrix}
a^c_{11} & a^c_{12} & . & a^c_{1q_1} \\
a^c_{21} & a^c_{22} & . & a^c_{2q_1} \\
. & . & . & . \\
. & . & . & . \\
a^c_{q21} & a^c_{q22} & . & a^c_{q2q_1}
\end{bmatrix} .
$$

4. Choice of q:

In order to suitably define the value of q, the energy retention capability of RFT is analyzed for a total of 80 (40 synthetic and 40 real-life pathological speech) signals with length N. Based on this investigation, it has been inferred that over 95% (statistical confidence ratio) of the signal energy is retained when q is reduced to $N/2$. Hence, during computation of RFT, q can be selected to be approximately half the size of row and column vectors respectively. This is done to minimize the number of non-dominant periodicities in the final matrix. That is, the maximum values for q_1 and q_2 are taken

as follows,

$$q_1^{max} = N_1/2 \text{ and } q_2^{max} = N_2/2.$$

The main objective of this step is to control the number of RFT points, defined by q_1 and q_2, during decomposition. This process aids in obtaining a reduced set of coefficients, in comparison to the FFT.

5. The next step involves identification of dominant coefficients in the computed signal-RFT matrix. This process indirectly aids in minimization of number of coefficients to be analyzed and in turn the overall dimensionality of the problem. To facilitate this, a ranking algorithm whereby a set of significant coefficients is extracted using the M best-basis algorithm. Here, M-best periodicities and their corresponding basis elements are extracted and the algorithm for this approach was previously proposed by Sethares et al [61]. In an effort to reduce the signal interaction terms, a soft threshold is applied on the RFT matrix. The threshold limit is chosen based on the 3σ energy distribution, such that the coefficients with values less than mean - 3σ are not included during the ranking stage.

4.5 Performance Validation Using Synthetic Example

In this section, a validation of the proposed RFT-based transform for AD analysis is carried out based on: (i) robustness of performance, (ii) verification of the correctness of the obtained periodicities, (iii) performance evaluation in comparison to the conventional FFT, (iv) investigation on the representation of signal (or auto-) and interfering (or cross-) terms in the intermediate TF plane (obtained using RFT), (v) potential for time-varying signal analysis, (vi) computational complexity and (vii) inherent sparsity. Each of these aspects are discussed in detail in the following subsections.

4.5.1 Robustness

In order to assess the properties of the RFT, the following simulated signal is considered:

$$x(n) = sin\left(2\pi\frac{n}{T_1}\right) + \sum_k \delta(n - kT_2),\ 0 < n \le N, \tag{4.18}$$

where n is the sample number, the period of sinusoid $(T_1) = 5$ and the period of impulse (T_2) = 7. Thus, the generated signal is composed of superposition of a sinusoid and an impulse train. The signal is corrupted using a white Gaussian noise. Figure 4.3 shows the FFT and RFT plots of the signal corrupted by white gaussian noise at 20 dB signal-to-noise ratio (SNR). In FFT, the sinusoid and the impulse are correctly captured i.e. the main peak is located at $26/128 = 0.203$ cycles/basis and the pulse train is captured at multiples of 0.14 cycles/basis. The periodicities of the signal components (peaks at $q = 5$ and 7) are equally well captured in the RFT plot.

However, when the SNR is decreased to 0.2 dB (Figure 4.4), the input appears completely distorted. In this case, although the FFT captures the sinusoid frequency, the spike train is not distinguishable from the many spurious frequency terms (with significant energy). In comparison to FFT, the RFT still manages to capture both the periodicities without any distortion. The FFT suffers from spurious signals and is not robust to presence of noise in the signal. These claims can be further validated from the results obtained from a noise-corrupted impulse train (with SNR = 20 dB), as shown in Figure 4.5.

4.5.2 Assessment of AD-RFT

In this evaluation, a sine wave input (period = 5) corrupted with Gaussian noise at 10 dB SNR is considered. Figure 4.6 show the AF, TF-FFT and TF-RFT plots for the sine signal. Here, the TF-FFT and TF-RFT coefficients are obtained by computing the FFT and RFT of Equation (4.14) and Equation (4.15) respectively. The performance validation is carried
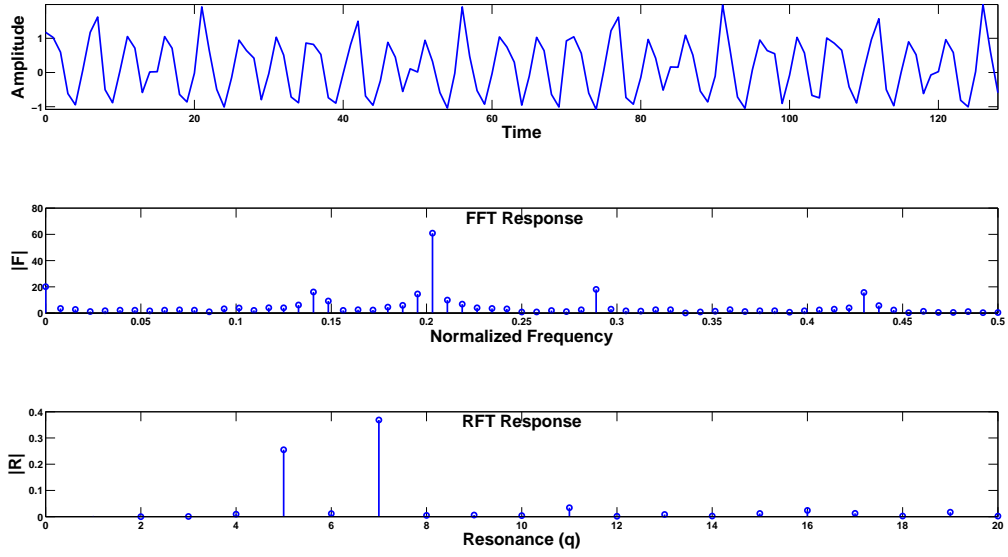
Figure 4.3: The FFT and RFT plots of an input consisting of: sinusoid (period = 5 samples) + impulse train (period = 7 samples) corrupted with white Gaussian noise (SNR = 20 dB)
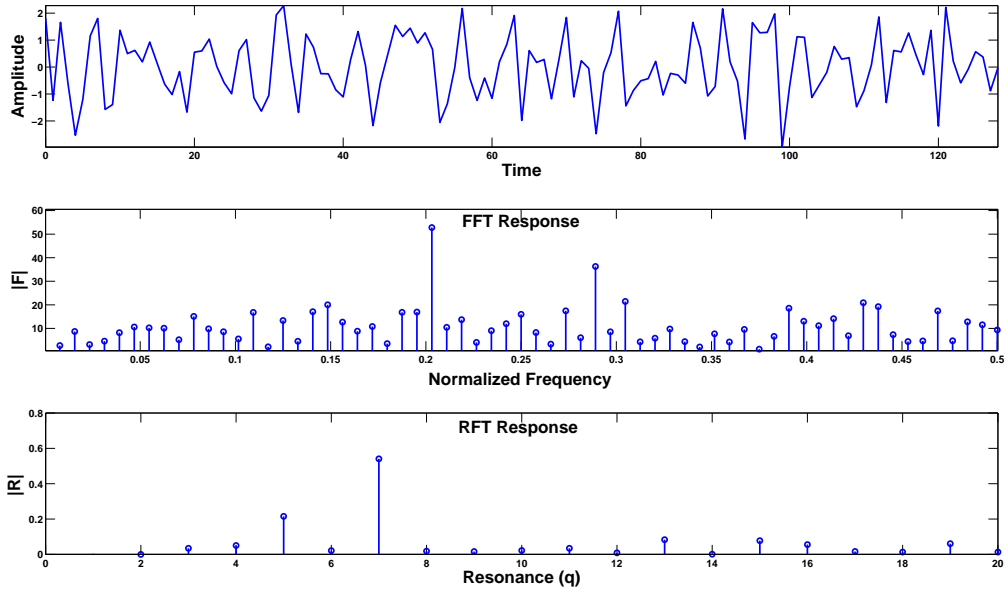
Figure 4.4: The FFT and RFT plots of an input consisting of: sinusoid (period = 5 samples) + impulse train (period = 7 samples) corrupted white Gaussian noise (SNR = 0.2 dB)

70
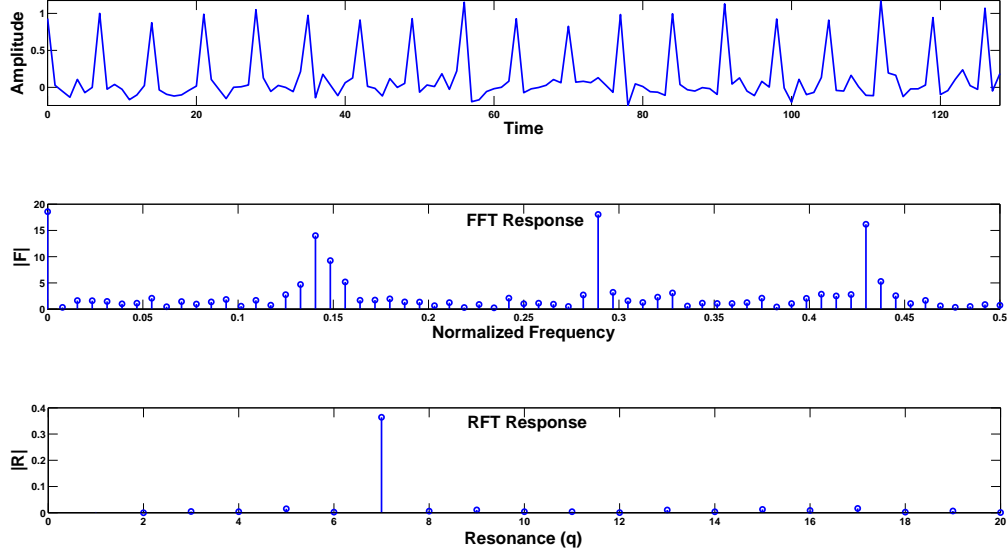
Figure 4.5: The FFT and RFT plots of an impulse train (period = 7 samples) corrupted white Gaussian noise (SNR = 20 dB)

out in the following three phases:

(i) *verify if the periodicity is captured:*

it can be noted (from the figure) that the dominant periodicity of the input is captured in the TF-RFT representation and this can further be validated from the corresponding TF-FFT plot. This result is expected since literature [47] suggests that characterization of stationary signals using RFT produces results similar to FFT.

(ii) *investigation of robustness in the presence of noise:*

Comparing the obtained 2-D plots for noisy sine input, it can be verified that in the presence of noise, the clarity of the TF-RFT representation is least affected in comparison to AF-based and TF-RFT plots.

(iii) *assessment in terms of dimensionality reduction:*

Figure. 4.7 shows the variation in reconstruction error, defined based on the ratio of the

71

energy of signal versus the TF-RFT coefficients, for different choices of q. It can be seen that, for lower values of q the number of co-resonance terms and in turn the leakage in signal energy is minimal, resulting in an approximation of the signal. A higher value of q provides optimal decomposition of the signal.

Further, in order to evaluate the performance for time-varying signals, a linearly decreasing chirp signal is analyzed. As shown in Figure 4.6 (g, h and i), it can be noted that the proposed TF-RFT approach, captures the inherent time-frequency connection using a fewer set of coefficients in comparison to the standard TF-FFT map. Further investigation using higher-order time-varying examples, can facilitate better understanding about the benefits and limitations of the proposed technique for time-frequency analysis.

4.5.3 Investigation of Auto- and Cross-terms

In this sub-section, the usability of RFT for TF analysis is investigated. Here, a two-component Gaussian atom is analyzed in the AD space using FFT (Eqn. 4.14) and RFT (Eqn. 4.15) respectively. Research shows that AD plays a significant role in time-varying signal estimation for both continuous and discrete time series.

For a multicomponent signal, the AD map consists of two terms: auto-terms and cross-terms. Auto-terms correspond to the dominant TF terms in a given signal. Owing to the bilinear nature of the transform, certain additional terms are generated during interaction between dominant frequency components. These are referred to as the cross-terms and are usually spread outside the origin in the AD-space. These interaction terms are usually tagged as being "unwanted" and not included in the analysis for NS signals [62]. Since the time-varying nature of the signal gives rise to these additional terms, characterization of these terms can be a novel directive in signal processing research.

Of the many available variants in TF distributions [14], different kernel filters are included in order to reduce the amount of interfering terms in the TF representation. The improvement
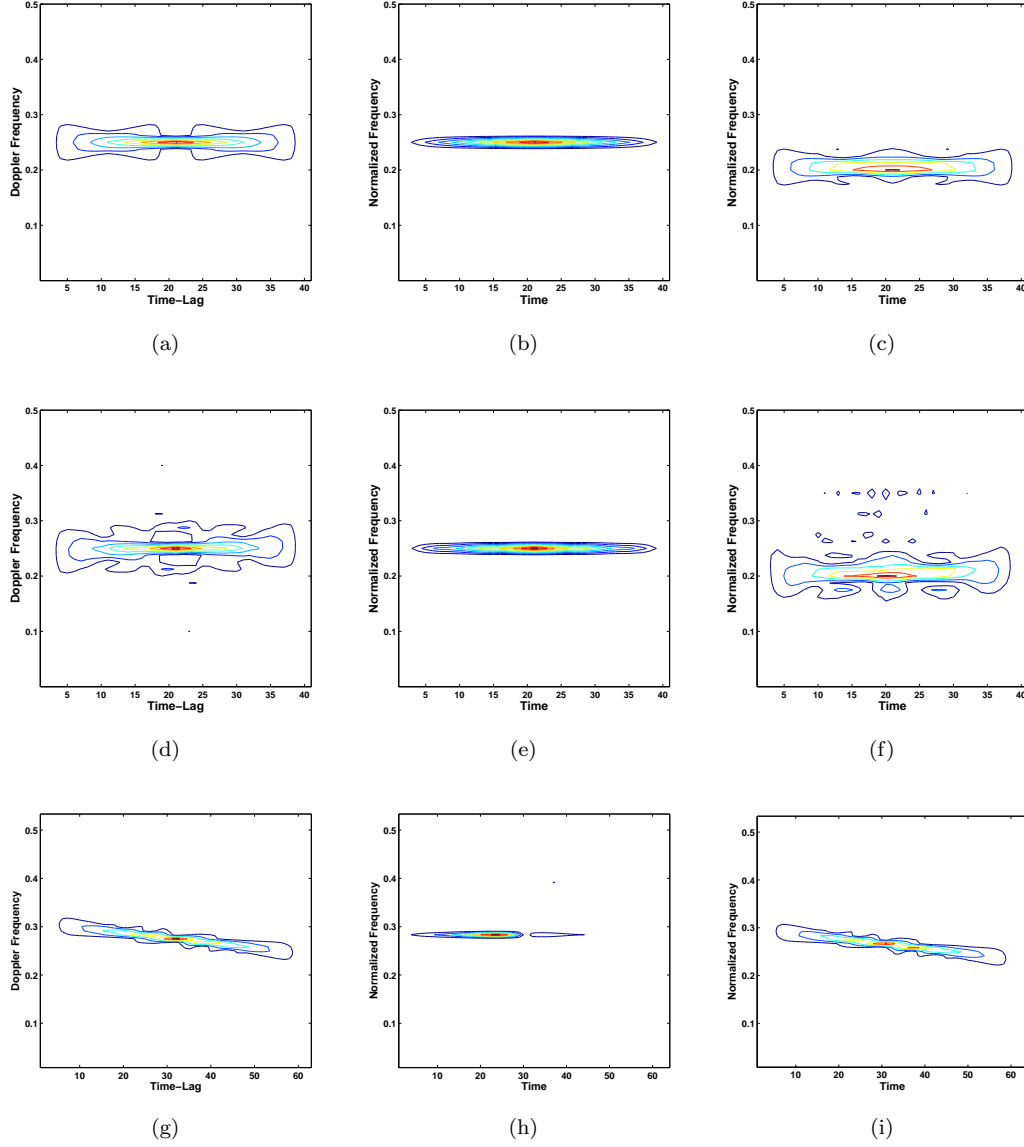
Figure 4.6: Representation of sine wave of period = 5 and chirp signal using AF, TFD-RFT and TFD-FFT respectively. (a), (b) and (c) correspond to the 2-D maps for the sine wave input without any added noise; (d), (e) and (f) correspond to the results obtained for a noisy sine wave signal (corrupted with white Gaussian noise signal for a SNR of 10 dB) and (g), (h) and (i) correspond to the chirp signal representations
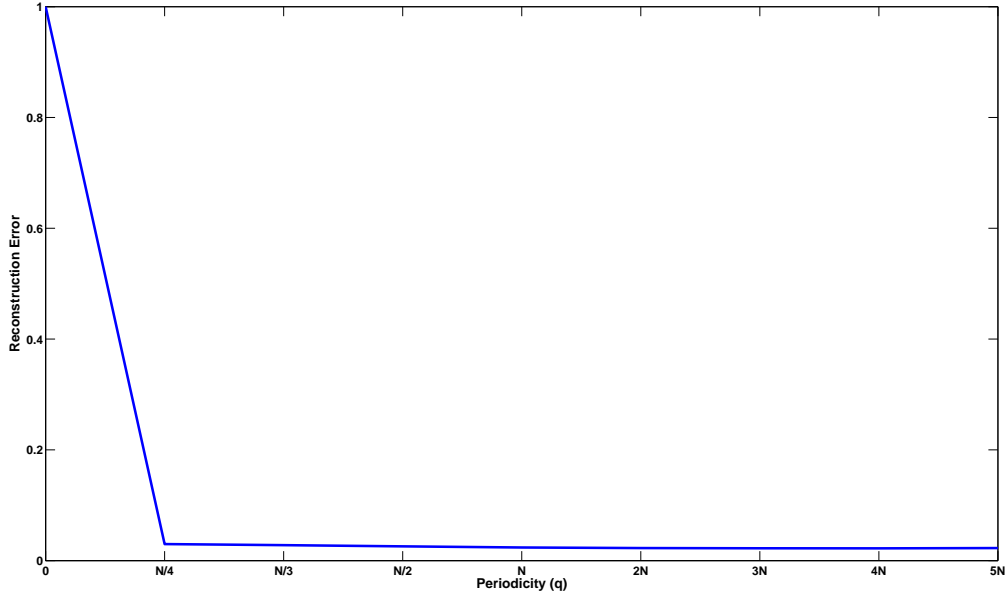
Figure 4.7: Variation in reconstruction error, defined based on the ratio of the energy of signal versus the RFT coefficients, for different choices of q

in the visual clarity in these techniques is achieved at the cost of reduced TF resolution i.e. smeared output results. In this work, a modified TF formulation obtained by computing the RFT of the time-varying autocorrelation function (i.e. a variant of the AF) is used for investigation of auto- and cross-term energy distribution.

Figure 4.10 shows the representation of a NS signal (Gaussian atom) obtained using the AD-FFT and AD-RFT formulations. From the plots, it can be verified that the AD-RFT plot retains reduced cross-term structure in comparison to the standard AD-FFT representation. Additional clarity can be obtained from Figure 4.10(c), where the overlap in the different regions (auto- and cross-terms) is better visualized. Further, it is interesting to note that the reduction is obtained without the loss in TF-resolution i.e. no smearing results. The reduction in interference can be owing to the computation of only co-prime resonance terms in RFT. In addition, a direct implication of this is by computing the absolute difference

74

between the AD-FFT and AD-RFT plots, more reasonable estimates of auto-term and cross-term energies can be obtained. These measures can in turn be used as features for designing characterization schemes for time-varying signals.

4.5.4   Spectral Leakage

Frequency-based transforms are considered efficient for assessment of a signal's spectral characteristics. The FFT (and its optimized implementation fast Fourier transform or FFT), one of the widely preferred methods for such analysis, provides accurate signal characterization when the measured signals are not limited in time. The DFT computes the spectral transform over a certain number of discrete frequency points called bins, while FFT can only be used with finite length signals whose length can be realized as a power of two.

The standard FFT suffers from spectral leakage loss i.e. presence of side lobes in addition to dominant frequency terms [56, 63]. This makes the larger signal component overshadow other significant small amplitude signals, thereby making them less detectable. Also, the dominant frequency terms no longer contain the complete energy, and there is an increase in noise and adjacent non-dominant frequency terms energy leading to reduced signal-to-noise ratio. Existing measures to mitigate the effects of signal leakage employ certain windowing techniques.

According to Shannon's sampling theorem [64], a continuous-time signal can be completely reconstructed from its discrete counterpart, if the highest signal frequency term is less than the sampling frequency i.e. the Nyquist rate. However, when the signal contains non-integer number of cycles, spectral leakage can still happen, and this results in energy spread from dominant component appearing across adjacent frequency bins. To avoid spectral smearing, smoothing windows are employed or an infinite length sequence is required.

This limitation of FFT can be owing to the fact that the Fourier coefficients are calculated at frequency components that are not harmonics of the fundamental component and is com-

puted for both even and odd multiples of the dominant frequency for finite length signals. In this work, an alternative directive to reduce the effect the spectral leakage is provided. The concept of RS can be suitably exploited in order to achieve the same. RFT, the modified Fourier representation obtained using the real sums of the nth powers of the qth roots of unity is used for this purpose.

The performance evaluation of RFT, to demonstrating its usefulness, is carried out using a sinusoidal signal example. The simulated signal is composed of a periodic component that repeats every fifteen samples. The chosen period is not a multiple of $1/N$ and hence spectral leakage can be better assessed. The normalized values of the FFT and RFT coefficients were computed and are used in the analysis. The validation is carried out for two sample datasets derived from the original data:

(a) S_1: sinusoidal signal considered for the entire length and

(b) S_2: randomly segmented signal whose length equals half the original data length, extracted from the 2048 length sinusoid.

Figures 4.8 and 4.9 show the FFT and RFT spectrum for each of the signals. It can be noted that, as the length of the considered signal is reduced, the signal energy concentration is disturbed and the overall clarity decreases. Here, for simplicity, the main lobe (contributed by dominant frequency component) is considered to be Gaussian-distributed, and define the signal spread factor as the ratio of the number of bins for the 3 dB energy to the total number of bins.

Comparing the spread ratios for the FFT and the RFT plots, it can be verified that the spectral leakage is more pronounced using the conventional FFT plots, while the RFT representation shows signs of negligible (spectral) leakage. The performance improvement obtained with the use of RFT can be owing to the use of co-prime resonance bins during signal projection in the spectral domain.

In addition, earlier investigation suggests the usefulness of RFT for capturing periodicities
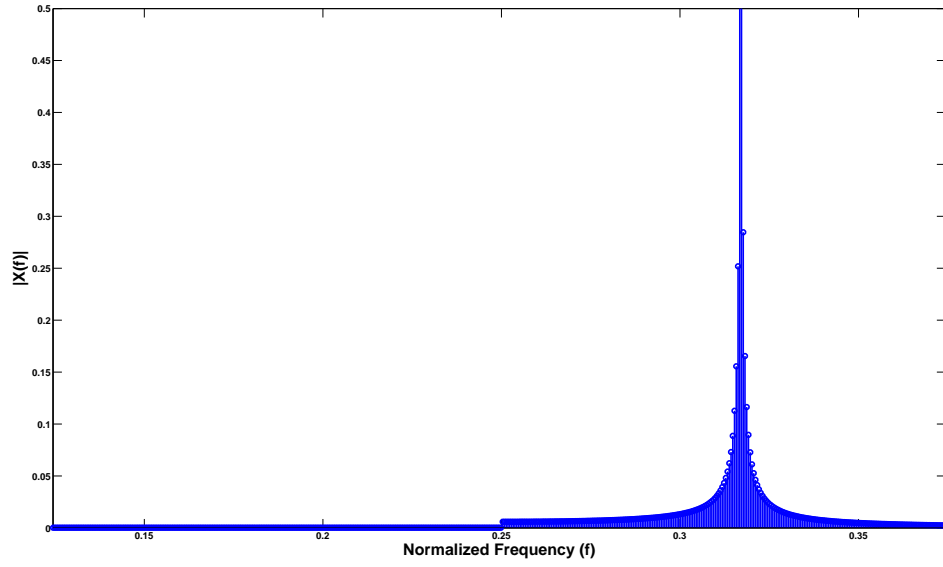
(a)
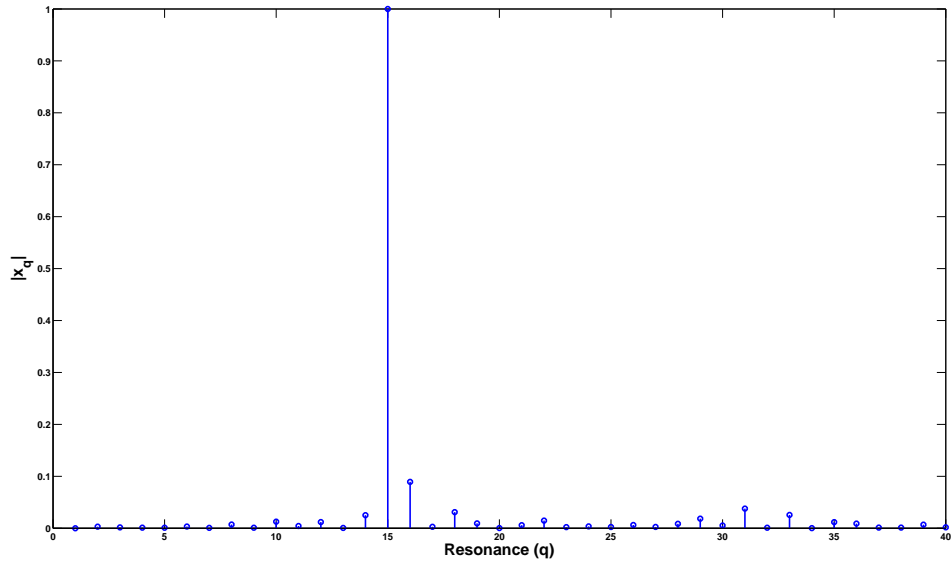
(b)

Figure 4.8: (a) FFT spectrum and (b) RFT spectrum of a sinusoid (length = 2048 samples), S_1, repeating after 15 samples
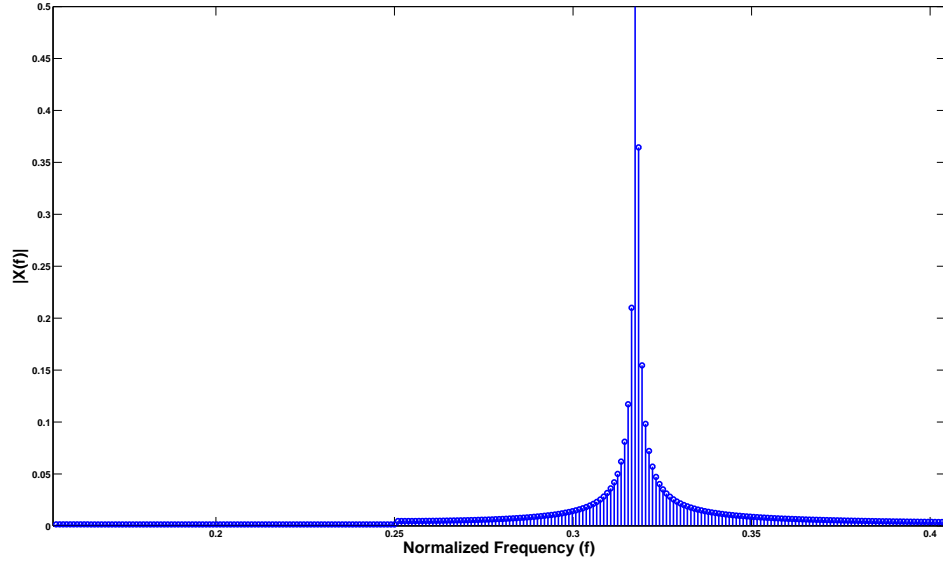
(a)
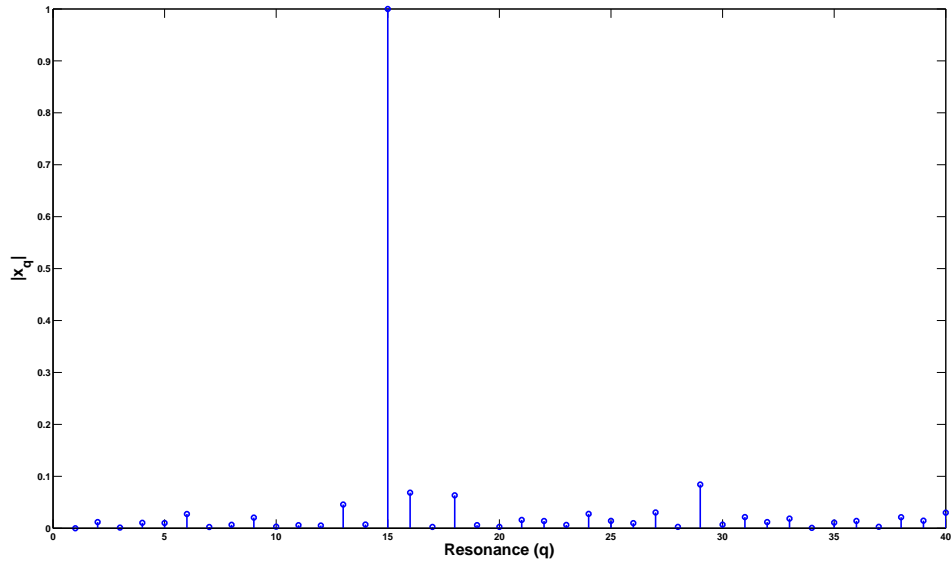
(b)

Figure 4.9: (a) FFT spectrum and (b) RFT spectrum of a sinusoid (length = 1024 samples), S_2, repeating after 15 samples
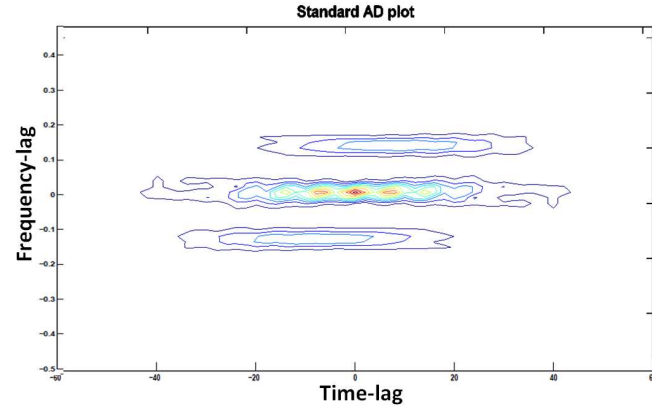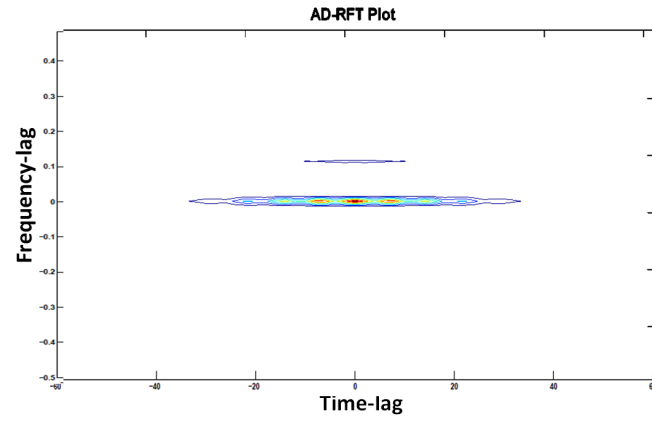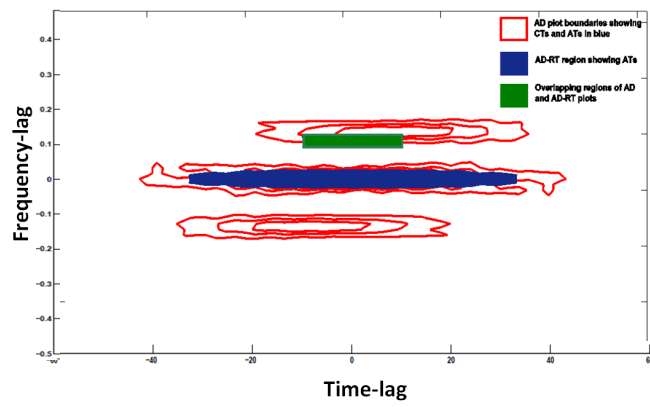
Figure 4.10: Representation of a two-component Gaussian signal in the (a) AD-FFT and (b) AD-RFT plane. Figure (c) shows the relative overlap of auto- and cross-terms in both the domains.

even when the signal is heavily corrupted by Gaussian noise terms [52]. The obtained results showcase the potential of RFT for robust analysis of certain signal types.

4.5.5 Computational Complexity

Further, the computational efficiency (defined in terms of the implementation complexity) of the proposed transformation is evaluated in an effort to evaluate the relative complexity reduction with reference to the standard DFT implementation. For a N-length signal, a typical DFT algorithm involves N^2 complex multiplications and $N(N-1)$ complex additions. In contrast to DFT implementation, the RFT uses integer computations and hence employs only shift and addition operations. Overall, the RFT computation involves N^2 shift and add operations only and no multiplications, thereby facilitating faster implementation.

Table 5.1 tabulates the theoretical complexity limits for each of the analysis, computed based on direct implementation of DFT. It can be noted that, since the calculations in RFT are made upon samples only; with respect to Moebius function, faster calculations can be achieved using the proposed technique. However, one of the current challenges in this technique is that, since this is one of the primary works exploring to usability of RFT in TFRs, the interpretability can be critical for more complex signals (like real-life signals) and hence needs to be investigated further.

4.5.6 Sparseness

The overall change in sparseness is evaluated for the conventional AD-representation obtained using DFT and the one using RFT with and without applying a soft threshold on the coefficients. This feature is quantified using a sparsity estimate proposed in [65]. For any

Table 4.1: Computational Complexity

Domain	Complexity	Comments
Input	$O(N)$	$N \rightarrow$ length of the signal
AD	$O(N^2)$	Owing to symmetric nature, can be reduced up to $O(N^2/2)$
DFT	$O(N^2)$	Fourier computation window length can be reduced based on the requirement
RFT	$O(Q^2)$	$Q = \gcd(q_r, q_c)$ and $Q < N$
TF-DFT	$O(N^4/2)$	
TF-RFT	$O(N^3 Q/2)$	

$p \geq 1$ (default '1'), p-sparsity measure is defined as,

$$S_p(x) = \left(\frac{1}{C_p}\right) \left(\frac{\left(\frac{1}{n-1}\right) \sum_{i=1}^{n} |x_i - m|^p}{\left(\frac{1}{n}\right) \sum_{i=1}^{n} |x_i|^p}\right), \tag{4.19}$$

where,

$$\left(\frac{1}{n}\right) \sum_{i=1}^{n} x_i \text{ and } C_p = \left(\frac{(n-1)^p + 1}{n^{p-1}}\right)^{\frac{1}{p}}.$$

A sparser vector will have a maximum value of S_p ($= 1$ for exactly one non-zero entry) and a redundant vector will have zero sparsity. Table 4.2 contains a preliminary comparison between sparseness estimates obtained for the synthetic NS signal data. It can be noted that, in comparison to the conventional AD-map, the RFT-based representation provides a relatively sparser representation. Also, by introducing a soft thresholding on the coefficients, the degree of sparsity can be substantially improved without the loss of critical information. Further, the overall complexity associated with using RFT, $O(q^2)$, is lesser in comparison to the standard DFT ($O(N^2)$), where $q << N$.

Table 4.2: Comparison of Sparseness Estimates for Synthetic NS Signal

	Overall Sparsity	
Technique	Before Thresholding	After Thresholding
AD using DFT	0.51	0.78
AD using RFT	0.75	1.00

4.6   Existing Challenges

1. Although RS can be used in extraction of certain relevant features from complex signals, the process of quantifying the obtained signatures into known arithmetic functions is still challenging. For instance, although the RFT signature for $1/f$ noise was identified for a phase locked loop, no plausible theory could be derived [47].

2. The choice of q is critical in RFT analysis. Hence, for random signal inputs, the selection criteria is still a heuristic approach and needs to be optimized further.

4.7   Chapter Summary

The concept of number theory has been revisited in recent times for signal processing. In this context, the signal processing benefits of the the integer-valued RS basis functions, obtained as the nth powers of the qth primitive roots of unity is presented. Further, a detailed investigation on the usability of RFT was carried out and certain significant properties of the 1-D Ramanujan-Fourier expansions were introduced.

A modification to the AF computation was proposed by changing the basis functions. Following this, a detailed validation of the proposed approach is presented using certain synthetic examples corrupted with white Gaussian noise. In addition, the spectral leakage behavior

in the conventional Fourier representation was compared and contrasted with the proposed transform. Finally, a formulation for the 2-D Ramanujan Fourier technique is proposed for multidimensional analysis.

This chapter introduced a novel class of transforms obtained by modifying the basis functions used during projection in the AD. The next chapter will introduce some novel variants to TF kernel functions using the AD representation.

Chapter 5

Methodology III: Kernel Design in AD

In recent times, for statistical knowledge discovery, kernel-based methods have attracted many researchers. Kernels are defined as the transformation function that computes a dot product in a feature space (i.e. Hilbert space). Kernels are designed to introduce useful non-linearities in order to facilitate handling of unusual input spaces such as complex NS signals. That is, kernel functions enable us to work in the feature space without the need to actually map the non-linearly classifiable data in higher dimensional space. This means that any linear algorithm where data can be represented as dot products can be made non-linear by substituting the kernel function in lieu of the dot product. In other words, kernels define the mapping of TF pairs in the signal by defining the distribution function. Hence, for obtaining new TFDs, their underlying kernel functions are modified.

In this chapter (Figure 5.1), some novel kernel functions are defined by modifying the characteristic function of a TFD. The proposed kernels are investigated with reference to cross-terms and TF resolution improvement.
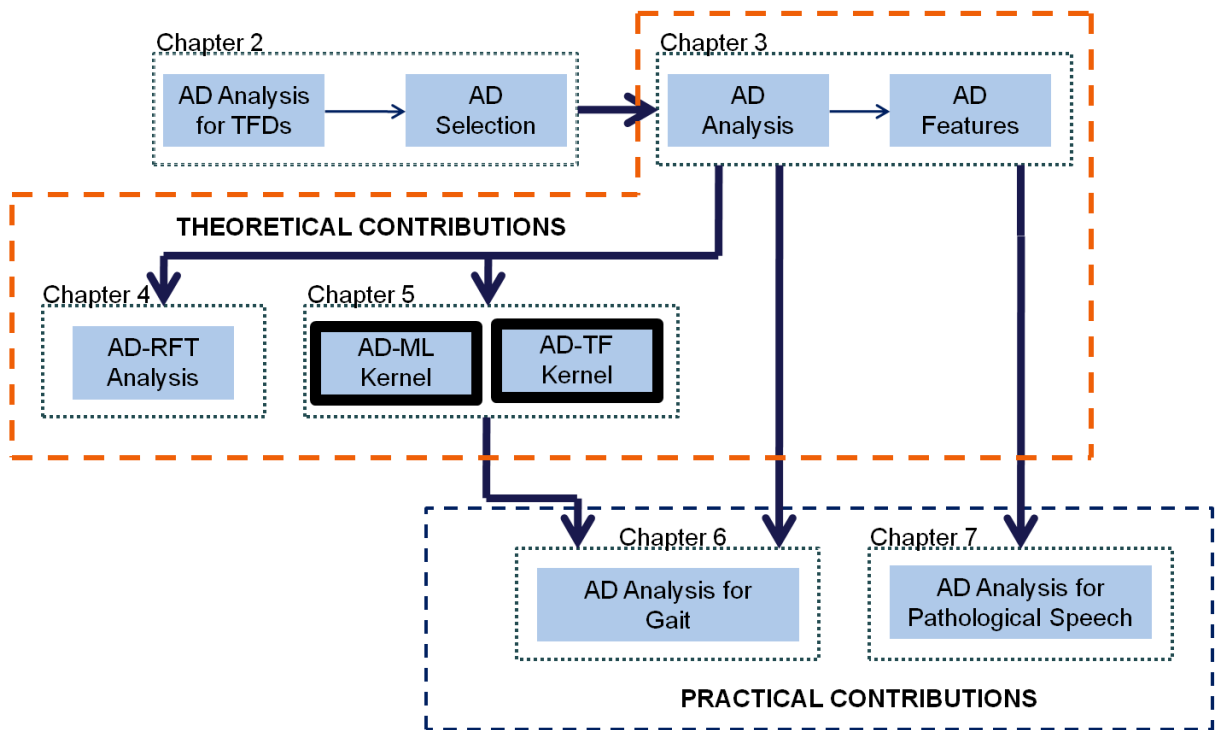
Figure 5.1: Chapter 5 - Kernel Design in AD

5.1 Kernel weighting functions

In this section, the Wigner-Ville Distribution (WVD) and some of the commonly used Cohen's class of kernels are explained. Following this, the proposed modifications to these standard kernels along with their performance will be discussed in detail.

5.1.1 Existing Kernels

Wigner-Ville Distribution

The bilinear distribution proposed by Wigner and Ville is defined as,

$$P_w(t,f) = \frac{1}{2\pi} \int x^*(t - 0.5\tau)x(t + 0.5\tau)e^{-tjf}d\tau. \tag{5.1}$$

In this distribution, the 2-D kernel function is unity i.e. $\phi(\theta, t) = 1$. This is a highly non-local distribution since the local and distant instants of time are weighed equally. Also, the values are always real even if the signal is complex. Since each part of the distribution is based on past and future values, when the signal is corrupted with noise the distribution possesses noise at those instants of time when it is not noticed in the input. There is also noticeable interference or cross terms.

Figure 5.2 shows a linear frequency modulated signal with 128 sample points and its TF representation using the WVD. Wigner-Ville kernel is one of the simplest kernels where the characteristic function is obtained directly from the AF and this function satisfies most of the essential mathematical properties. However, the kernel used here is an all-pass filter and hence suffers from presence of non-negligible cross-terms that affects the visualization of the signal components which can be further verified from the figure.

The corresponding AD-equivalent representation is obtained using two methods:
(1) the AF can be computed directly from the characteristics function and
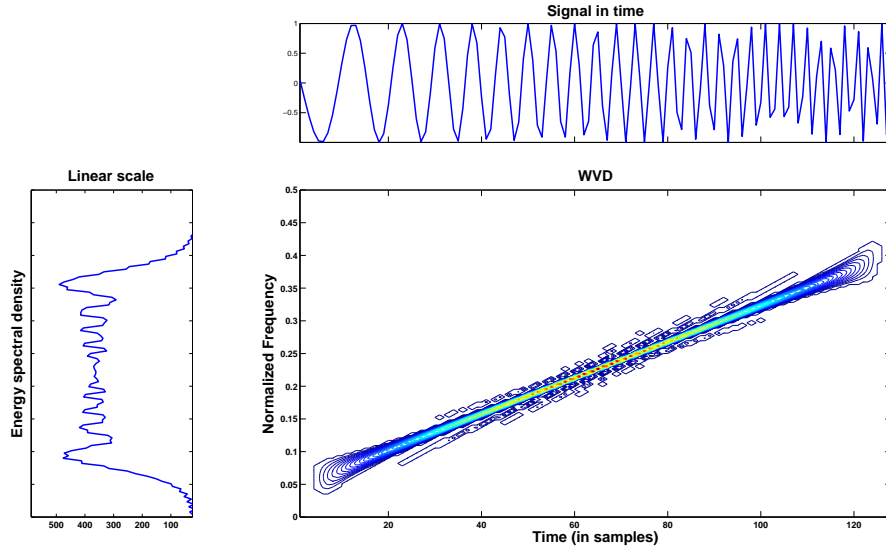
Figure 5.2: Linearly frequency modulated signal input (length = 128 samples) and the corresponding TFD

(2) by reverse mapping the TFD.

Figures 5.3 (a) and (b) represent the AD representation plots obtained through direct computation from Eqn. 3.8 and by reverse mapping from the corresponding TF representation respectively. The obtained plots are compared in terms of the difference in energy concentration in the representation. Both the plots appear similar, and in terms of the representation discrepancy, a negligible error of 2.8 X 10^{-14} results. This indicates that the reverse mapping method generates reasonably accurate AD-map. In addition, such a method facilitates ease of computation of the AF matrix for kernel-based TFDs (other than the all-pass kernel).

For further analysis, on other available kernels, a Gaussian or Gabor atom input is used. The input signal structure is as shown in Figure 5.4. The signal is considered because of the fact that Gabor atoms have good energy compaction in the TF plane and the corresponding analysis could provide additional insight about this domain with reference to signal representation. The Gaussian atom used in this analysis, consists of 2 dominant frequency terms. The TF and the AD plot for the Gabor atom obtained using WVD are shown in Figure

87
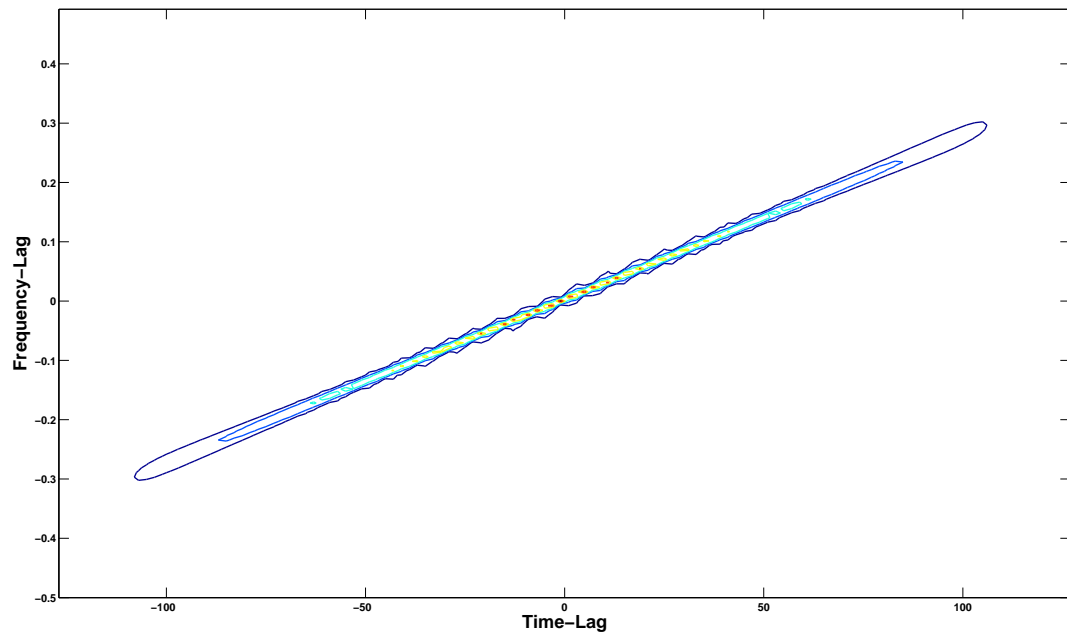
(a)
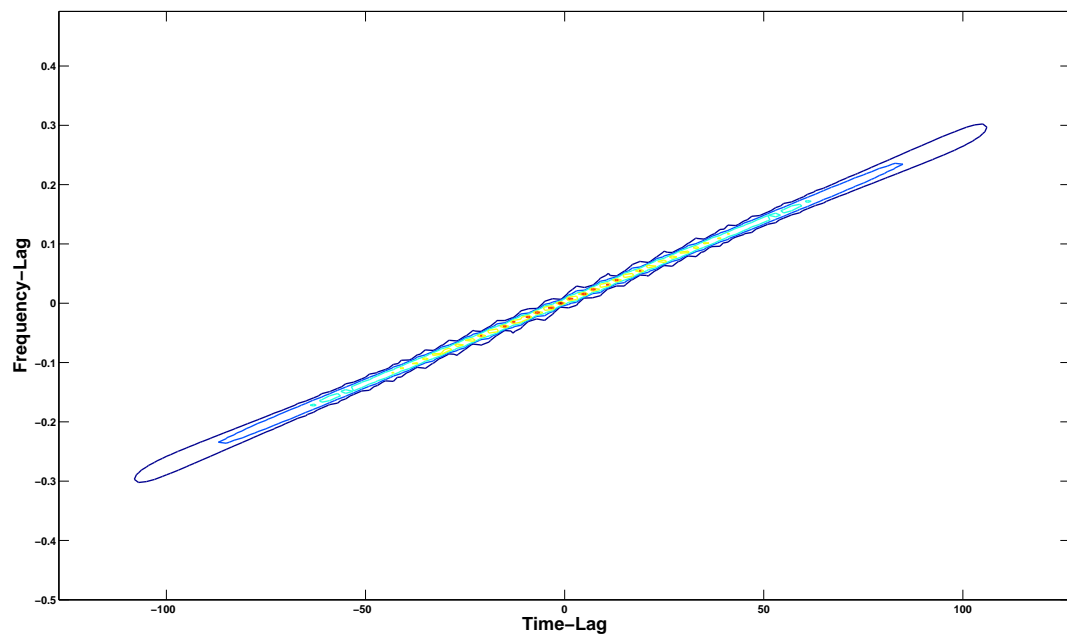
(b)

Figure 5.3: AD plot obtained through (a) Direct computation and (b) through reverse mapping from TF representation

88
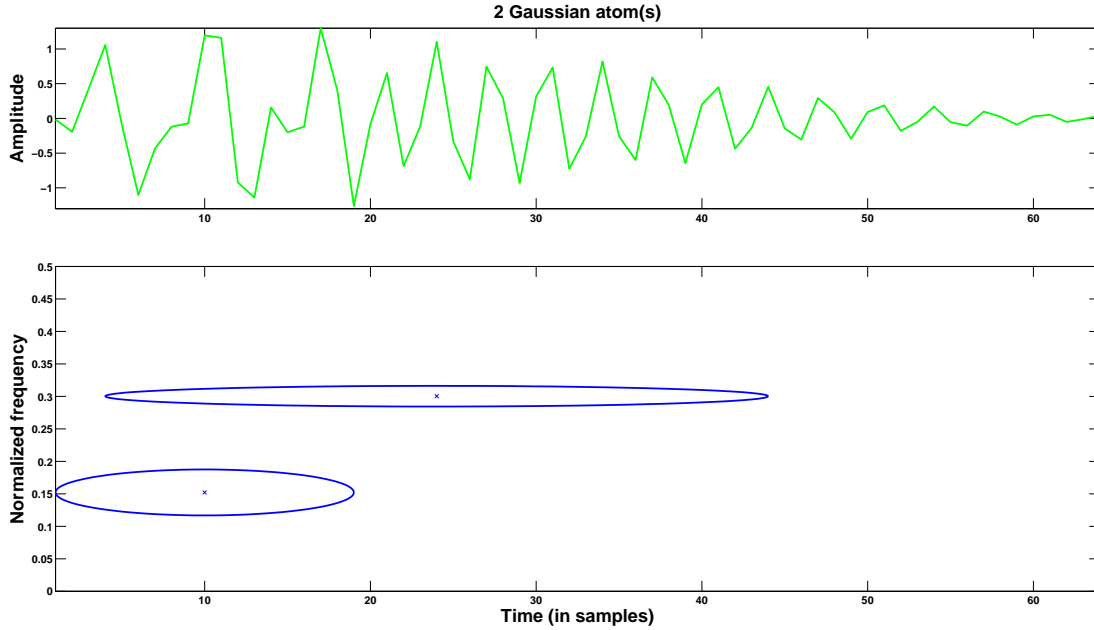
Figure 5.4: Gaussian atom input with two dominant frequency terms

5.5. It is evident from the properties and the obtained results that WVD has excellent time and frequency resolution but generates cross-terms when analyzing multi-component signals. These cross-terms (or artifacts) means that the distribution shows energy which does not really exist at particular t, f coordinates. Further research in kernel design is targeted towards obtaining reduced cross-terms in the representation. The so-obtained distributions are referred to as the smoothed distributions. Some of the representations will be discussed in the following subsection.

Choi-Williams Distribution

The Choi-Williams Distribution (CWD) is a TFD that retains many desirable properties and yet attains a partial attenuation of cross-terms. The characteristic function for this
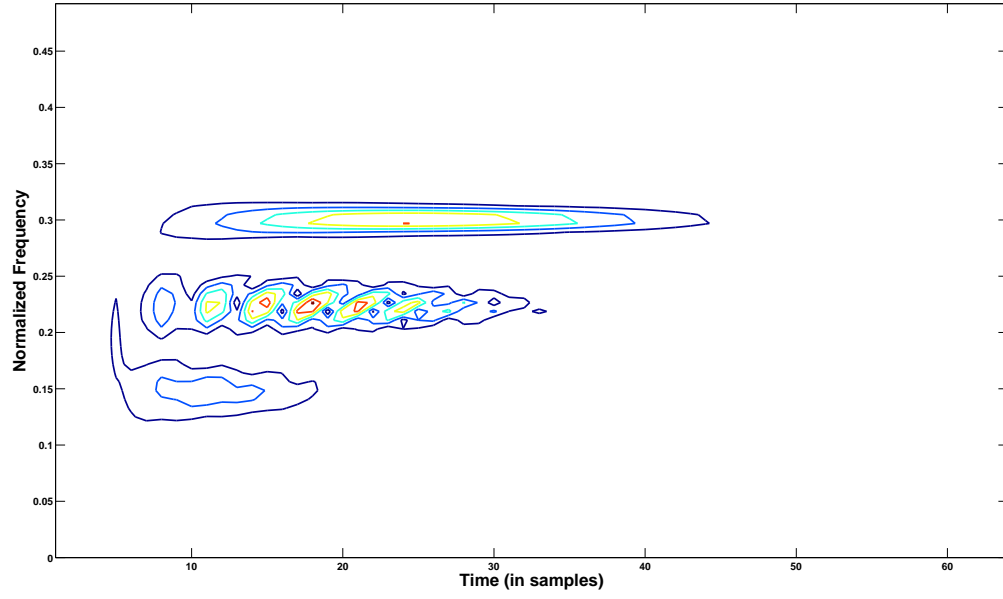
(a) TFD
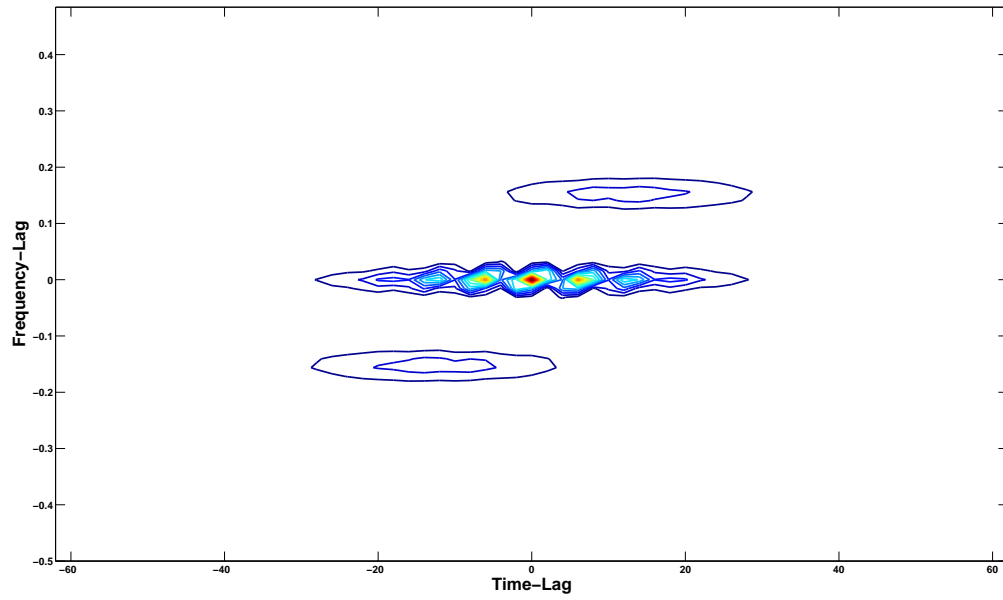
(b) AD

Figure 5.5: TF representation and AD-plot of the Gaussian atom using WVD

90

distribution is defined by the following kernel,

$$\phi(\theta, t) = e^{-\theta^2 t^2/\sigma}. \tag{5.2}$$

The CWD uses an exponential kernel. Figure 5.6 shows the TF and AD representations respectively. It can be noted that the cross-terms are attenuated by use of this kernel; however, this comes with a price of added smoothing in the representation. Here, the σ parameter of the exponential kernel can be suitably varied to generate narrow and wide-band representations. Most frequently, σ is chosen to be unity and a value of less than one is used for narrow-band signal selection and to obtain maximum cross-term reduction.

Margenau-Hill Distribution

The Margenau-Hill distribution (MHD) is obtained as the real part of Rihaczek distribution. In a Rihaczek distribution, the interaction energy between the components of a signal is restricted to an infinitesimal interval, and the corresponding kernel is an infinitesimal bandpass filter which can be approximated by the following expression,

$$\phi(\theta, t) = e^{j\pi\theta t}. \tag{5.3}$$

The corresponding MHD kernel is defined as,

$$\phi(\theta, t) = \cos(\pi\theta t). \tag{5.4}$$

This kernel is chosen because of the fact that, MHD satisfies most of the properties that are listed for WVD, except for modulations and unitarity. The response obtained by using the Margenau-Hill kernel on a Gaussian atom input is shown in Figure 5.7. Here, both the auto-terms and the cross-terms appear smeared out in both the TF and the ambiguity plots.
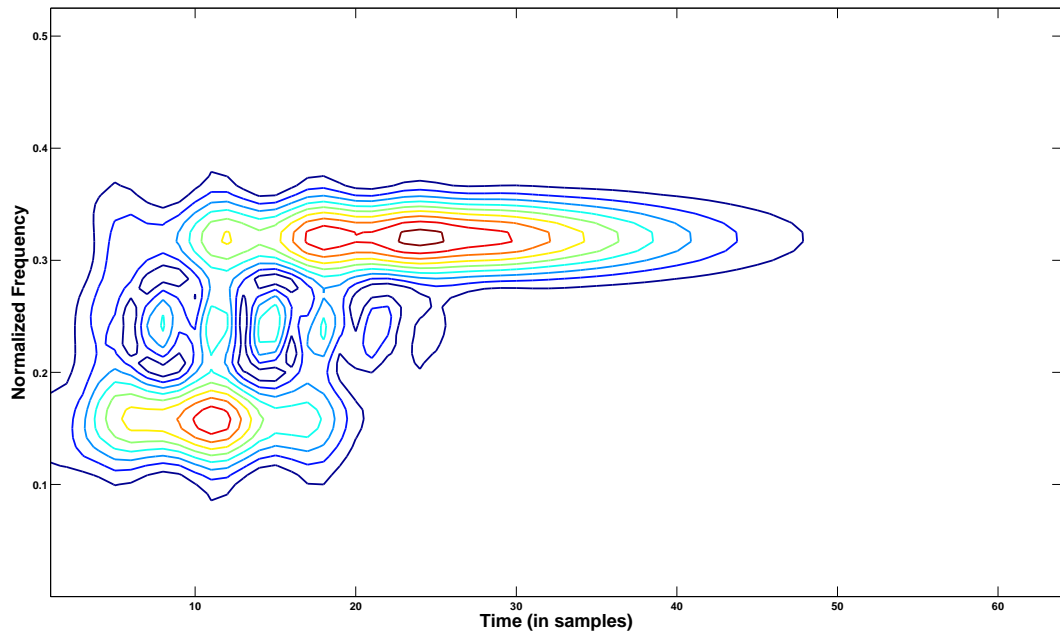
(a) TFD
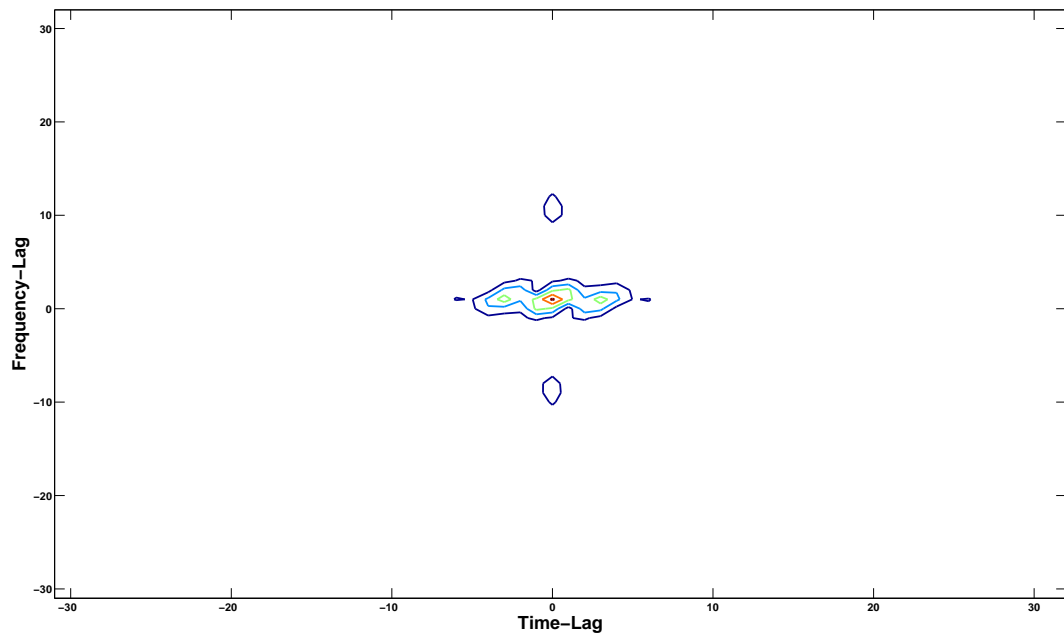
(b) AD

Figure 5.6: TF representation and AD-plot of the Gaussian atom using CWD

The cross-terms appear prominently in the representation. The main difference between this and the cross-terms that appear in a WVD is that, in WVD the cross-terms appear in between the frequency terms and in Margenau-Hill distribution, the cross-terms appear at the same time instant as a signal term but at a different frequency. The use of the Rihaczek (or Margenau-Hill) distribution for signals composed of multi-components located at the same position in time or in frequency is not generally advisable, since the interference terms will most likely be superposed to the signal terms for multicomponent signals, when the components are close to each other. Similar results were obtained for the case of Page distribution as well. The other kernel definitions (Page, RID, etc) similar to CWD and MHD, are smoothed distributions, hence are not discussed here. For details regarding the kernels, refer to [14]. In the following sub-sections, the proposed kernel variants in AD will be discussed.

5.1.2 Proposed Kernel Modifications

For any given signal, $x(t)$, in order to facilitate efficient processing, there exists a need to employ specific distribution functions such that accurate concentration of signal results along both the time- and the frequency-axis i.e. an inverse function of the group delay gives the value of the instantaneous frequency term. However, in practice, especially when the signal has multiple frequency components, it then becomes difficult to define the ideal distribution function. Hence, in most cases, the distribution is assumed to be the equivalent to the sum of individual component distributions.

TFDs have been widely recognized as a powerful tool for analysis of time-varying multicomponent signal types. However, owing to the inherently generated bilinear products, these distributions suffer from presence of cross-terms. The problem associated with kernel design tradeoff for cross-term suppression and TF resolution (and localization) is addressed in this work. A kernel basically performs an arbitrary TF weighting or mask the signal terms, for
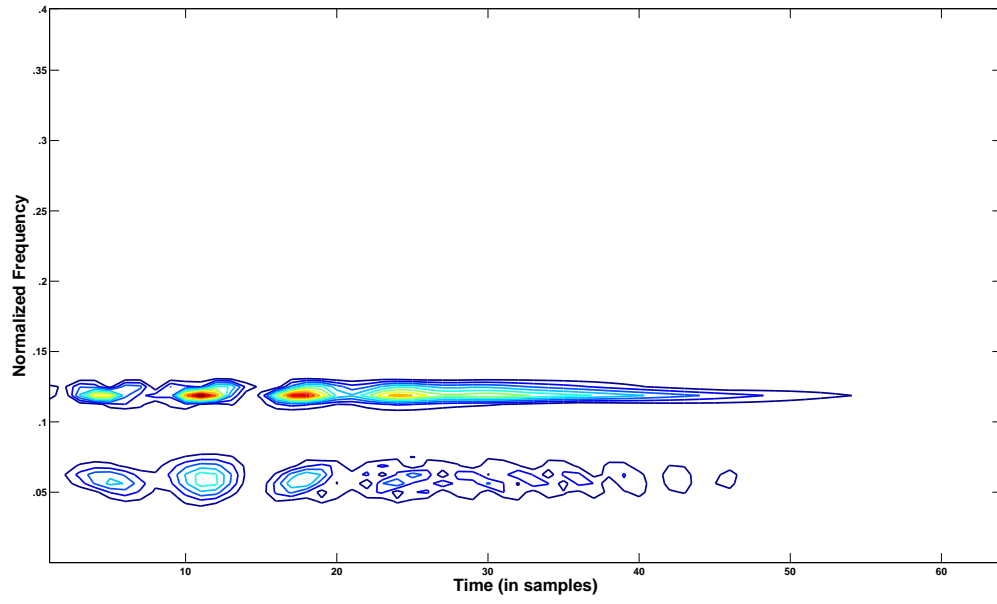
(a) TFD
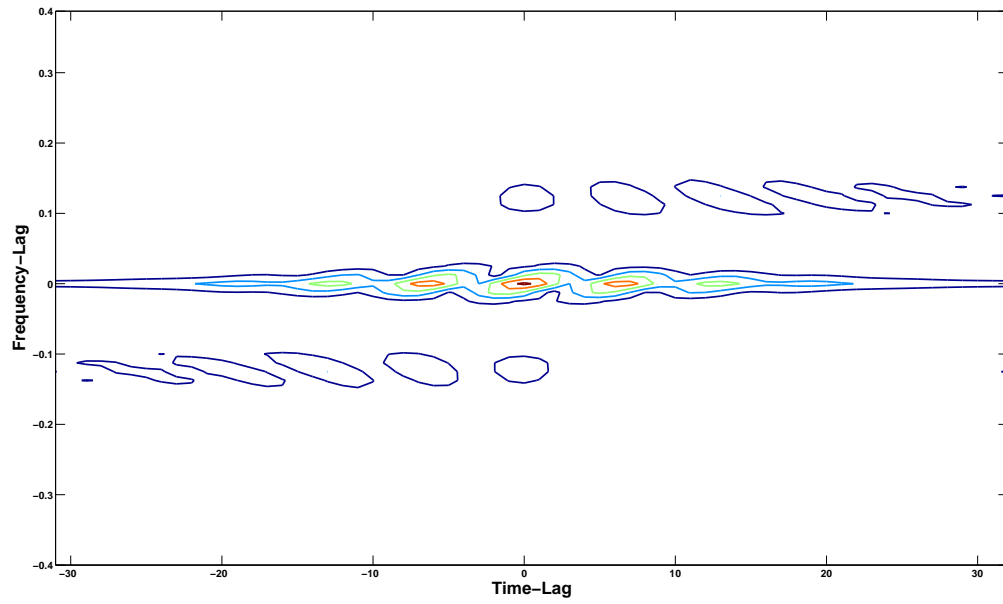
(b) AD

Figure 5.7: TF representation and AD-plot of the Gaussian atom using MHD

a predefined combination of TF pairs, from attenuation.

The key TF distribution properties are: non-negativity, time-shift, frequency-shift, time and frequency marginals, instantaneous time and frequency computation and reduced cross-terms. The challenge lies in the fact that time and frequency support and certain other properties cannot be easily enforced for AD-based design. For any discrete time signal, the AF is continuous and periodic about frequency lag and discrete in time-lag axis. Hence the TF kernel should exhibit similar characteristics.

A variety of different kernels were analyzed and the interpretable results were discussed in the previous section and in Chapter 2. Based on our investigation, it has been verified that, in order to facilitate characterization of a time-varying signal, energy distribution across cross-terms should be considered a discriminative feature as it varies between different signals. Hence, the chosen kernel should not be devoid of cross-terms along with providing a reasonable TF localization. In this work, two specific modifications to the existing TF kernels (namely Choi-Williams and Margenau-Hill kernels) are discussed in the following subsections.

Scalogram-Modified Choi-Williams Distribution

It is interesting to note that, Choi-Williams kernel exhibits rapidly temporal decaying behavior for smaller lags, thereby approximately satisfying the time support property in the ambiguity plane, making it one of the suitable candidates to start with for kernel design. Also, this kernel satisfies some of the essential TF constraints. In the ambiguity plane, the Choi-Williams kernel is given as,

$$\phi(\theta, t) = e^{-\theta^2 t^2 \sigma}. \tag{5.5}$$

The equivalent in the TF plane is,

$$\phi(\tau = 0, f) = \delta(f), \tag{5.6}$$

$$\text{and } \phi(\tau \neq 0, f) = \frac{1}{\sqrt{\frac{4\pi\tau^2}{\sigma}}} e^{\left(-\frac{f^2}{4\tau^2/\sigma}\right)}. \tag{5.7}$$

In the modified version, a Morlet wavelet function is used as the time and frequency smoothing filters instead of the default window functions. The complex Morlet wavelet filter (bandwidth: f_b and centre frequency: f_c) is defined as follows,

$$\psi(x) = \frac{1}{\sqrt{\pi f_b}} e^{i2\pi f_c x} e^{-\frac{x^2}{f_b}}. \tag{5.8}$$

This filter is used to wade out the cross-terms along the time-lag and the frequency-lag axis. The corresponding response is as shown in Figure 5.8. The smeared cross-term component that was prominent for the case of CWD is greatly reduced and an overall enhancement in the visualization quality is obtained. Additional constraints on kernel width is imposed based on the signal size and the number of frequency bins are then modified in proportion, to achieve better tradeoff between cross-term reduction and TF localization. The obtained kernel also satisfies the marginal properties. It is interesting to note that, the pass band of the defining kernel results in a narrowed signal representation in the ambiguity plane. When mapped on to the TF plane, there is a corresponding widening of the representation in the TF plane, since the kernel convolution function at each lag resembles a Gaussian.

Modified Margenau-Hill Distribution

The second contribution in kernel modifications for obtaining good TF resolution along with a reasonable reduction in cross-terms exploits the inherent characteristics of MH kernels. The MHD is one of the generalized classes of Cohen's distributions since it is deduced from linear TFDs. This is one of the distributions that satisfy most of the key properties. However,
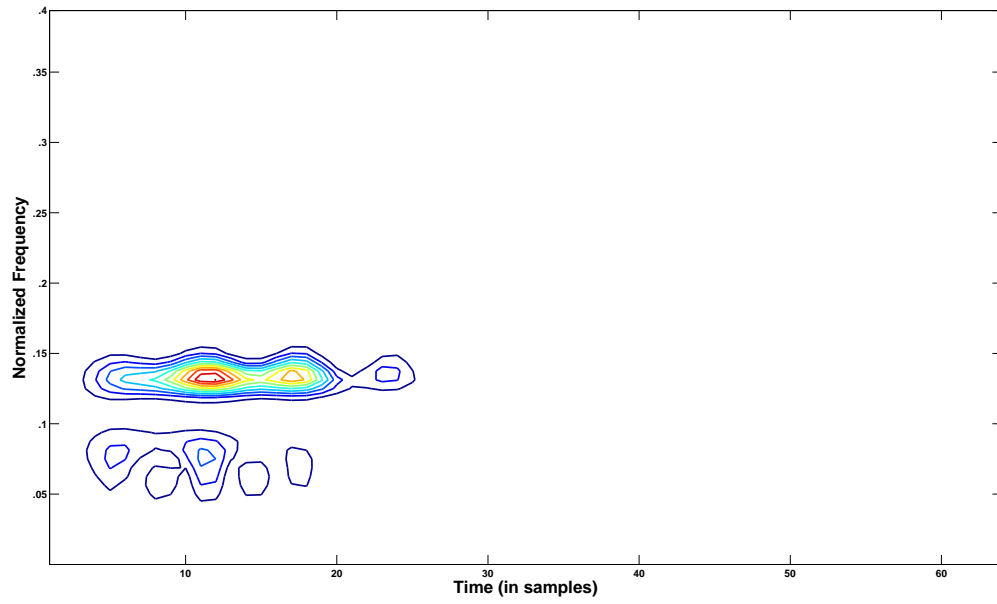
(a) TFD
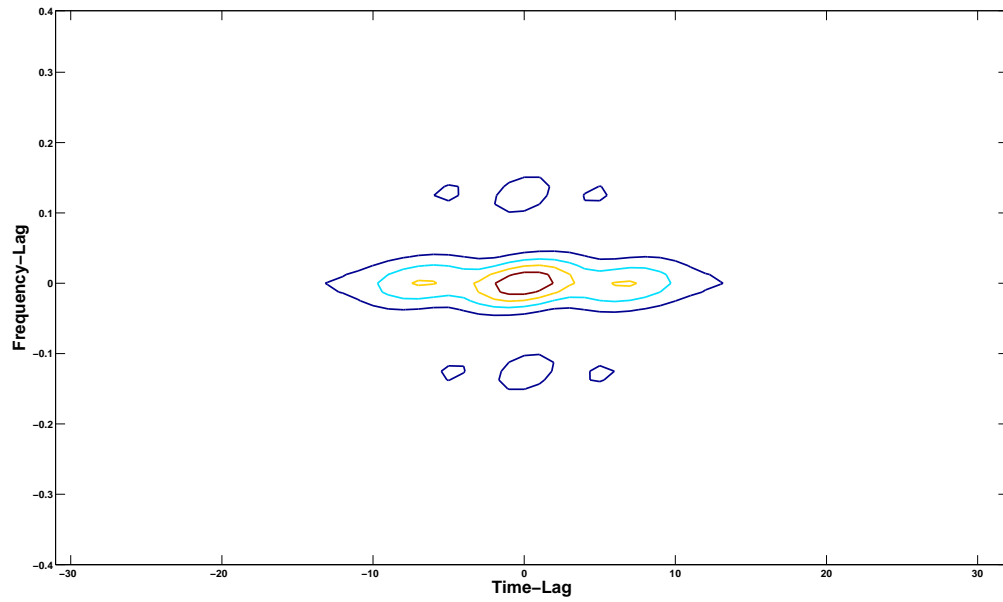
(b) AD

Figure 5.8: Top: TF representation and Bottom: AD-plot of the Gaussian atom using modified Choi-Williams kernel

the fact that the signal and interfering terms appear at the same time-instant cannot be ignored and can be crucial in computation of relative cross-term energy. Unlike Wigner-Ville kernel, where the cross-terms are unmanageable, the specific placement of cross-terms in this distribution is exploited to develop enhanced kernel types.

In order to achieve this, a spectrogram-based modification is performed on the MHD. This process reduces the smoothing factor thereby resulting in visibly less cross-terms in the AD and this can be further verified from Figure 5.9. In MHD, the kernel characteristics corresponding to desirable TF properties are well established [66]. Analysis of different kernels and their modifications to obtain new representations are discussed in [40, 67, 68]. This concept is extended for modifying the MH kernel for obtaining better visualization and characterization of signals.

Since the MHD is derived from linear functions, there is a good possibility that cross-terms can be better resolved and or reduced when the characterizing kernel function is suitably selected. Here, the kernel uses a squared version of the cosine function, which is then transformed using the Fourier function to obtain the distribution. A magnitude squaring involves computation of the inner product between the kernel function and the signal component, followed by squaring. This calculation is chosen because it is a linear operation $(\cos^2\theta = (1 + \cos 2\theta)/2)$ and hence does not introduce any additional cross-terms. Such a product kernel can also facilitate separable smoothing along the time- and the frequency-axis. Various linear modifications of the kernel, such as kernel ratio, and sum of two kernels, were analyzed and a squaring function is selected.

$$\phi_{MMH}(\theta, t) = (\delta(\theta, t) + \cos(2\pi\theta t))/2. \tag{5.9}$$

Except for bilinearity, all of the properties of the MH distribution are retained using the modified kernel. The proposed kernel function satisfies non-negativity, time and frequency
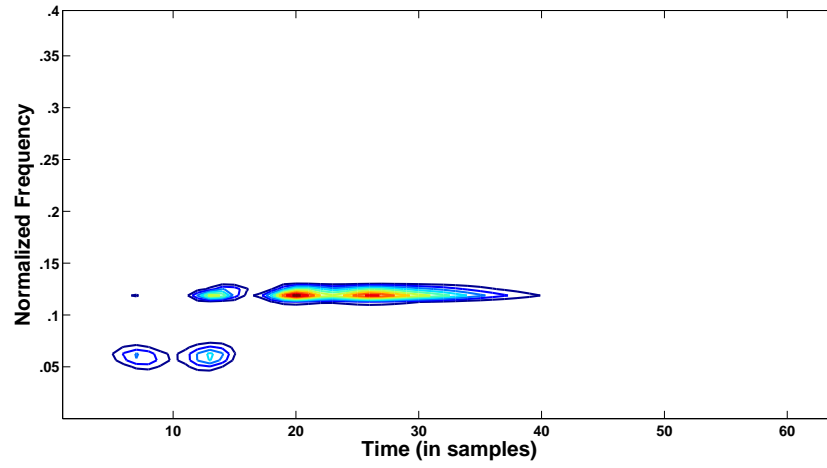
(a) TFD
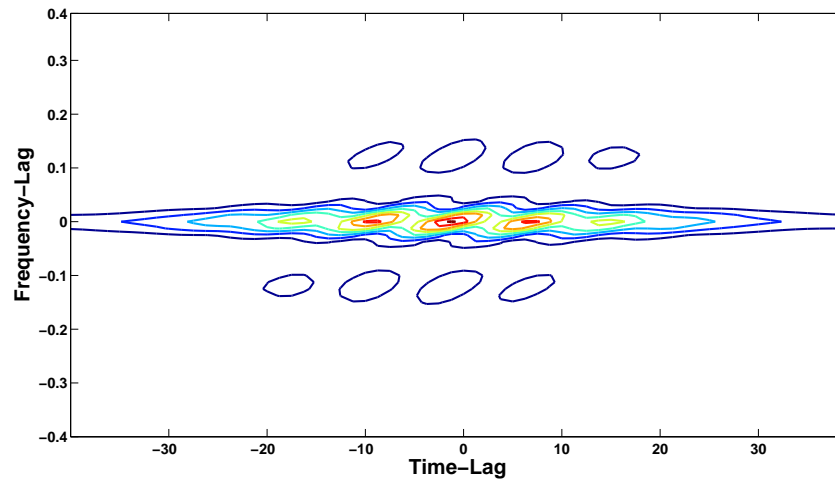
(b) AD

Figure 5.9: TF representation and AD-plot of the Gaussian atom using spectrogram modified Margenau-Hill kernel

99

invariance and energy conservation properties. The TFD and the AD plots are shown in Figure 5.10. In comparison to the MH plot in Figure 5.10 (b), it can be verified that this kernel results in an overall reduction in smoothing of auto-terms and also a substantial reduction in the amount of the cross-term concentration. In the TFD response, it can be seen that when compared to the original MH plot, the signal interfering cross-terms are greatly reduced; however the individual frequency component clarity which was previously visible in the original MH is not clear with the current effort to reduce the cross-terms. Owing to this, the suggested modification is not suitable for multicomponent analysis and necessitates further investigation on MH-based kernel design.

5.1.3 Applications of Proposed Kernel Modifications

Cross-Term Reduction:

Although cross-term reduction is not the main objective of the kernel modifications, this application is provided to show that TF resolution improved and the suggested kernel does not smear the output.

In the AD space, it is known that the auto-terms are concentrated around the origin and cross-terms are placed further away from the origin. Having said this, one way to reduce the interfering signal content is by use of appropriate 2-D low pass kernel functions. Figure 5.11 shows the AD response characteristics before and after the application of a low pass filter respectively. A simple example whereby, a low-pass filter designed with smooth transition characteristics is used to separate out the cross-terms completely, leaving out only the auto-terms. This is usually employed in applications such as, blind source separation and doppler spectrum estimation [41].

Signal Classification:

In order to validate the discrimination capabilities of the kernels, 2 non-stationary signals are used : speech and gait. A detailed investigation on the suitability of this method is
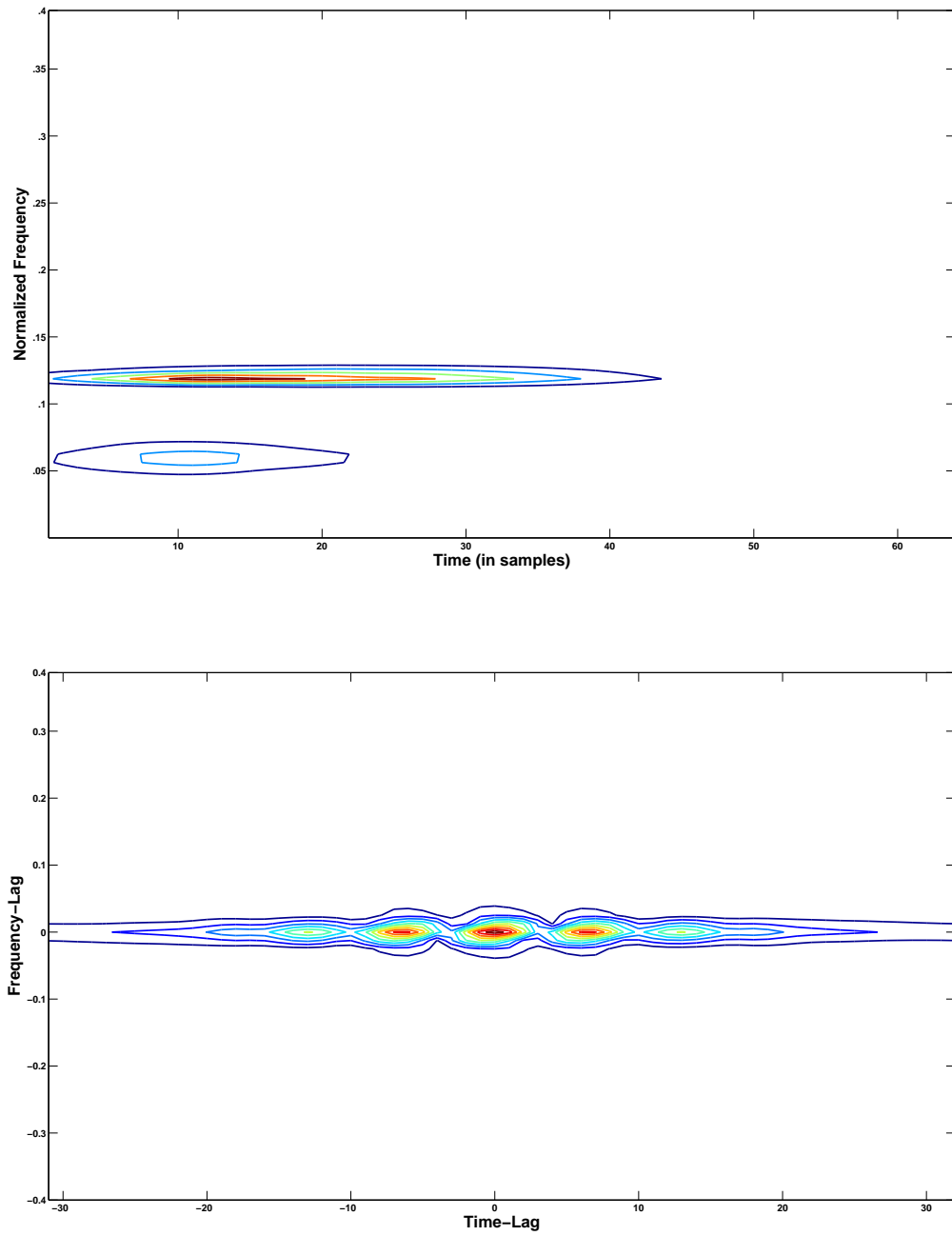
Figure 5.10: Left: TF representation and Right: AD-plot of the Gaussian atom using modified version of Margenau-Hill kernel

(a) Before Filtering
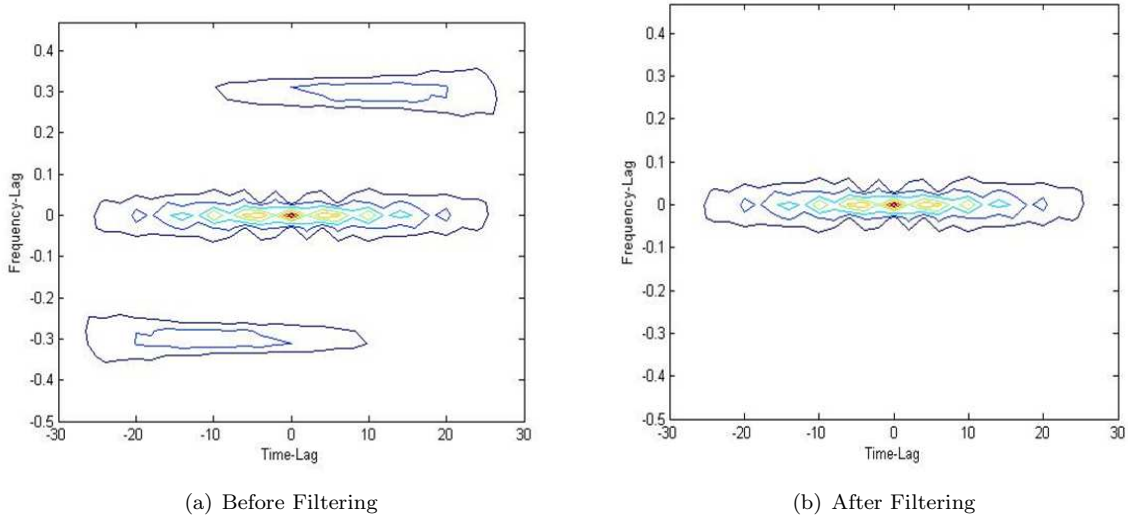
(b) After Filtering

Figure 5.11: AD plots of Gaussian atom before and after low-pass filtering using kernels for cross-term reduction application

presented in Chapters 6 and 7.

From Figure 5.10, it is evident that the proposed kernel functions provides a visually less smeared representation of auto-term energy patterns and thereby a corresponding reduction in cross-term concentration. Further, to facilitate classification between NS classes, signal-specific feature vectors can be suitably extracted from RDRs. These features along with class labels can then be used in integration with a linear classifier to obtain signal discrimination.

5.2   Kernel Design

The conventional approach for NS signal classification is as shown in Figure 5.12. A given NS input signal, $X : x(t)$ is first transformed into a higher-dimensional space using time-, frequency- or TF analysis tools. From the mapped space, representative features are extracted for classification. For NS inputs, owing to non-availability of discriminative feature definitions that suits all available signal classes, robust non-linear classifiers are often used to facilitate design of high-performance schemes. The function of the non-linear classifiers is to

transform the input (can be feature) space in a higher order Hilbert space, where the intra class (within a class) variance is minimized and the inter class (compared to all given classes) is maximized. The objective is to characterize (and categorize) different signals (from the same class) using a linear decision boundary.

These high-performance classifiers employ self-learning kernel functions that transform the feature space into a more separable map. To achieve this purpose, the input space is mapped on to the Hilbert space and are assigned specific weight vectors. Following this, certain optimization criteria is included in order to adaptively update the weight vectors till best discrimination is obtained.

In this context, kernel-based methods, like support vector machines (SVM), have drawn increased interest in the signal processing community, as they allow a simpler implementation of nonlinear algorithms, for the case of complex real-life signals. These algorithms were initially used to solve binary classification problems [69]. Later, owing to the attractiveness that stems from the strategy (kernel trick) used, extensive research [70, 71] in this domain extended the feasibility of such approaches for solving multiclass nonlinear problems that requires higher (or in some cases infinite) dimension processing.
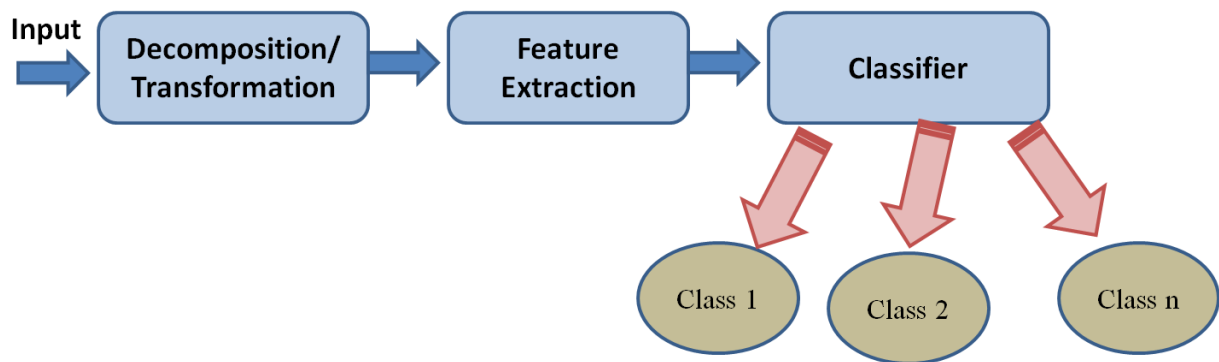
Figure 5.12: Block Diagram for Discriminative Kernel Learning in AD

5.3 Discriminative Kernel Learning in AD

5.3.1 Research Focus

Real-life signals (such as biomedical) are generated from stochastic systems and hence show certain trends that govern their underlying NS time-varying behavior. Investigation shows that there exists a need for an automatic recognition system for such signals, in order to facilitate an optimal characterization for discerning between different classes of data.

It is known that for statistical knowledge discovery, kernel-based methods have attracted many researchers since these functions have the ability to provide a linear or a non-linear map for most signals. From a signal processing perspective, there can be at least two distinct kernel design motives:

(i) k_{tf} - kernel(s) that act like a filter and which then defines a quadratic TFD. The main purpose of these kernels is to provide a suitable representative space for maximally discriminative feature vectors

(ii) K_{ML} - kernel(s) that are employed in a classifier with an ability to iteratively update the governing parameters in order to achieve higher distance of separability between different classes of signals.

In most available works, both these kernels are used extensively and independently. The main objective of this research work is to efficiently design a TF kernel in the AD, by exploiting the underlying principles of the machine learning kernels, in an attempt to reduce the number of computation steps needed to deduce an optimal feature space, and in turn to obtain a linear separation among different classes.

5.3.2 TF Learning Machines

In recent years, TF and time-scale distributions were designed using kernel-target alignment strategy, whereby the TF kernels undergo self-learning with respect to the task. Honeine et al [72] introduced the concept of TF learning machines, in which the TF reproducing kernels are aligned using a kernel-target alignment algorithm. The so-obtained kernel Gram matrix alignment is then treated as a maximization problem with an application specific cost function. The performance benefits offered by this new class of functions for supervised and unsupervised learning problems were later investigated and reported in [73, 74]. Very few works in literature [75, 76] have targeted integration of TF kernels and machine learning for signal analysis.

Since the choice of relevant TF features is still an open-ended question, this approach for obtaining kernel modifications has been explored. Analysis on existing TFDs indicates that their characterizing kernels need to satisfy some significant properties to facilitate TF analysis. These properties are investigated from a machine learning (ML) kernel perspective and are presented below.

- **Non-Negativity**:

  Most available ML kernels can be represented as symmetric functions and satisfy Mercer's theorem [77]. A kernel is said to be positive-definite if and only if for all non-zero vectors Z, Mercer's theorem is satisfied and for any n x n matrix, M,

$$Real(Z^* \, M \, Z) > 0, \tag{5.10}$$

  where Z^* is the conjugate transpose of Z.

  The AD is a correlative domain and kernel mapping in this domain results in a symmetric representation on both ends of the TF plane. Also, the generated set of cross-terms

is placed away from the origin. Here, the number of non-zero elements depends on the amount of frequency interaction (cross-) terms for each signal. Since the mapped coefficients exhibit symmetry, from 5.3.2,

$$R(Z_{AF}^* \, D_{AF} \, Z_{AF}) > 0. \tag{5.11}$$

Here, D_{AF} refers to the determinant of the signal coefficients matrix obtained using AF. Hence, if a kernel is positive-definite, the TFD representation results in non-negative values. Also, it is interesting to note that, WVD (the only distribution that does not satisfy non-negativity), can be made positive-definite by convolution with phase-space Gaussian function. Similar modifications can make other available negative ML kernels usable in constructing TFDs.

- **Time and Frequency Marginals**:
  Most existing ML kernel functions do not provide correct marginals. For applications that require correct marginals, the spectrogram-based ($K_{S_{WB}}$: wide-band and $K_{S_{NB}}$:narrow band) kernels can be used to cancel out the TF trade-off and the resulting kernel is then be said to be possess pseudo-optimum marginals. Mathematically, a given kernel, K, is modified as,

$$K' = K + K_{S_{WB}} + K_{S_{NB}}. \tag{5.12}$$

- **Finite Support**: Positivity and (almost) correct marginals together guarantee strong finite support [14]. That is, distribution is zero for all TF instants when the signal does not exist. Also, since a combination of narrow-band and wide-band spectrogram is used in modifying the kernel, the generated kernel offers better cross-term suppression properties and a good TF resolution.

- **Translation Invariance**:

  A positive semi-definite kernel can be made translation invariant by multiplying it with vanishing moments such that the function is radially non-increasing and is bounded.

5.3.3 Adaptive Kernel Learning in AD

The proposed discriminative kernel design is intended to serve two main objectives:

1. to efficiently incorporate the non-linear mapping during the TF representation stage using the concept of kernel machines, so as to facilitate extraction of robust features that are separable using a linear classifier and

2. to investigate the usability of cross-term (or interfering terms) components, when a signal is bilinearly transformed into the TF domain, by building discriminative feature space.

A general block diagram of the approach is shown in Figure 5.13. Here, the signal is first transformed into the AD-space and the mapped coefficients are masked selectively using a discriminative or a generative (machine learning) kernel function. A discriminative kernel allows sampling of the data conditional to the observed variable and hence, offers limited performance benefits for differentially variable classes of inputs, while the kernels obtained using generative models (like Fisher kernel obtained using Gaussian mixture model, etc) exploit the input characteristics during characterization.

The kernel masked coefficients are then used to define the feature vectors and class labels. For signal discrimination applications, this feature space is then used in integration with a suitable module (such as pattern classifier, hypothesis testing, and $k-$means analysis). Despite the conceptual simplicity and available advances in reproducible kernels, very few works [78, 79] that associate TF analysis with kernel machines are available. From these existing works [78, 79], it can be verified that this area of information extraction has not yet

been fully exploited by TF schemes.

In this research, a characterization scheme for solving binary (2-class) and multiclass prob-
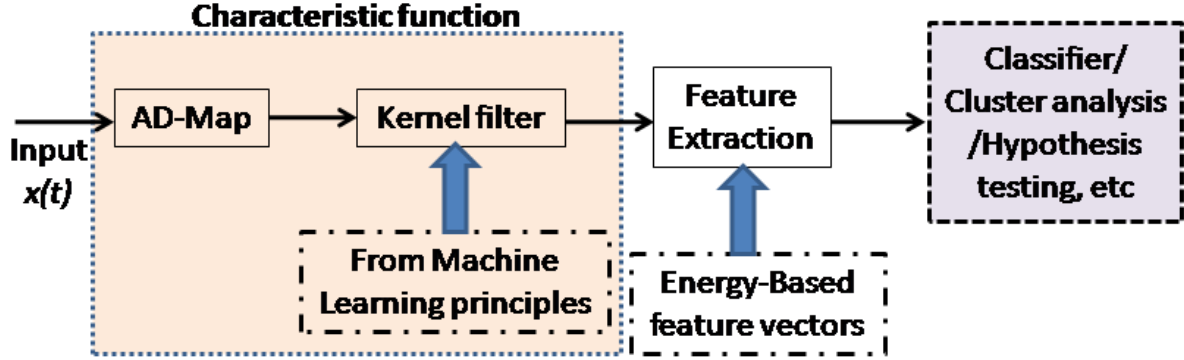
Figure 5.13: Block Diagram for Discriminative Kernel Learning in AD

lems associated with non-stationary signals is proposed. Here, the signal transformation along with kernel mapping is analyzed in the AD. The peculiarity of AD is that for multicomponent signals, the auto- (or signal) terms are localized around the origin and the generated cross- (or interfering) terms are positioned away from the origin [14]. Such a representation can facilitate characterization of auto- and cross-terms relatively easier than in the TF space. Also, AD is defined using the AF, which computes the time-varying autocorrelation for a given signal, and hence can be suitably used to capture the short-term and long-term correlations for a time-series signal.

Computation of AD-Fisher learning machines

Consider a set of n signals x_1, $x_2,...,x_n$, each of length d_i, where $i = 1, 2, .., n$. Each of these signals are then represented in AD (X_k's) according to their Wigner-Ville distributions. These coefficients are then mapped on the higher dimensional space using a suitable machine learning kernel. In our analysis, the Fisher kernel [80] is used owing to the following unique characteristics:

(i) these kernels compute the log-probabilistic model of the signal, which is then used along with the signal itself during the higher-dimensional mapping, thereby providing a suitable

characterization map and

(ii) are able to process variable length datasets.

The Fisher kernel is defined by the following equation [81].

$$K(X_i, X_j) = U_{X_i}^T I^{-1} U_{X_j}, \tag{5.13}$$

where, I is the identity matrix and the Fisher score matrix is obtained using expectation maximization algorithm [82]. This ensures that the expectation of the AF-transformed input is maximized for obtaining good discrimination. Mathematically, the kernel is defined as [80],

$$U_X = \bigtriangledown_\theta \log P(X|\theta). \tag{5.14}$$

That is, the log probabilistic model of the AF (that characterizes the signal behavior in the correlation domain) is used as the kernel function for higher order mapping. Following this, the final step in transformation involves computation of the Gram matrix for the kernel mapped coefficients. That is, for a given set of vectors, the Gram matrix is computed as the Hermitian matrix of inner products. In machine learning context, the use of Gram matrix ensures that the resulting kernel is positive semi-definite [83].

The Gramian Fisher kernel transformed coefficients are then used for investigation of distribution of signal- and cross- terms in the AD-space. Although most research in TF attempts to suppress the cross-terms for better visualization [14, 84], these terms cannot be completely ignored as they tend to carry some information about the interactions among signal frequency components and their removal also results in increased smoothing of auto-terms, thereby a reduction of TF resolution in the distribution. In this work, distinct features that characterizes the generated set of signal- and interfering-terms in AD were extracted, thereby providing a new direction for quantifying complex NS signals. More details about the feature extraction process is provided in the next subsection.

5.3.4 Feature Extraction

In order to validate the usability of the proposed technique for signal characterization, novel characteristic feature sets are extracted from the kernel transformed coefficients. Five AD-specific feature vectors were extracted from the mapped space (AD). The computed features are energy based and are used to characterize the signal. The relative energy distribution between auto- and cross-terms is analyzed for obtaining suitable metrics for discrimination between classes. Owing to the symmetric nature of this representation, the dimensionality of analysis space automatically reduces from O(N X N) to O(N/2 X N).

(i) *AD - distribution threshold (f_1)*: For signals with minimal number of frequency interactions, the region corresponding to auto- and cross-term overlap can be visualized in AD. This region is computed by dividing the AD-space into proportional bands based on relative energy distribution. However, for more complex signals this region is not easily perceivable and in addition to this region, the overlap between cross-terms also adds on the distortion in representation.

Here, in an effort to investigate cross-terms, f_1 or the maximum relative auto-term energy (within the band where the signal energy drops to certain threshold before increasing again slightly to represent the interfering term concentration) is computed. Our analysis also indicate that AD regions corresponding to relative energy of (about) 70% or less, are a result of overlap of cross-terms and signal terms.

(ii) *Maximum Energy (f_2)*:

$$f_2 = \left((AD_{Fisher})^2_{real} + (AD_{Fisher})^2_{imag} \right)_{max}. \tag{5.15}$$

(iii) *Relative energy across cross-terms (f_3)*:

$$\text{Total Energy } (TE) = \sum_{|AD_{Fisher}| \geq AD_{threshold}} |AD_{Fisher}|^2, \tag{5.16}$$

$$f_3 = TE - |(AD_{Fisher})_{within\ f_1}|^2. \tag{5.17}$$

Here, $AD_{threshold}$ is introduced in order to reduce the nearly-zero coefficients during the analysis. (iv) *Fisher score (f_4)* [80]: For a given signals' kernel-mapped AD-coefficients, the Fisher score is computed using the log-likelihood of the probabilistic model and is defined in terms of $P(X|\theta)$ (Eqn. 5.14). Here, the probabilistic model is calculated based on the AD-representation of the signal. As a pre-computation step, a selective weighting is used whereby the coefficients within the band specified by f_1 are given less weightage compared to those that falls outside this band, so that a detailed cross-term analysis can be facilitated. (v) *Sparsity factor (f_4)* [65]: The sparsity factor is introduced in the feature space, in order to investigate the change in sparseness among the different classes in AD and to test if they contain any clue that can in turn facilitate the signal characterization. For any $p \geq 1$ (default '1'), p-sparsity measure is defined as,

$$S_p(x) = \left(\frac{1}{C_p}\right) \left(\frac{\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}|x_i - m|^p}{\left(\frac{1}{n}\right)\sum_{i=1}^{n}|x_i|^p}\right), \tag{5.18}$$

where,

$$\left(\frac{1}{n}\right)\sum_{i=1}^{n}x_i \text{ and } C_p = \left(\frac{(n-1)^p + 1}{n^{p-1}}\right)^{\frac{1}{p}}.$$

A sparser vector will have a maximum value of S_p (= 1 for exactly one non-zero entry) and a redundant vector will have zero sparsity.

Algorithm:

S-1: Compute the AF for the given time-varying input.

S-2: Compute the Gram matrix, so that the matrix is positive-definite.

S-3: Map the AD signals in the higher dimensional space using the Fisher kernel.

S-4: Extract AD-specific feature vectors, as defined in the previous subsection.

S-5: The feature space, defined using the extracted feature vectors and specific class labels is then used for quantitative assessment of the proposed algorithm, using a linear discriminant analysis (LDA) classifier.

5.3.5 Validation

Dataset

The triangular database originally examined in Breiman [85] is considered in order to evaluate the performance of the proposed algorithm. Saito [86] later extended the length of the dataset (from 21 to 32 for each signal), so that the defined signals are of dyadic length. In this dataset, the three classes are formed from a convex linear combination of shifted triangular wave inputs defined by h_1, h_2 and h_3 functions, making this one of the challenging discrimination problems. The corresponding characteristic equations are,

$$\textbf{Class 1: } x^{(1)}(i) = uh_1(i) + (1 - u)h_2(i) + \epsilon_i, \tag{5.19}$$

$$\textbf{Class 2: } x^{(2)}(i) = uh_1(i) + (1 - u)h_3(i) + \epsilon_i, \tag{5.20}$$

$$\textbf{Class 3: } x^{(3)}(i) = uh_2(i) + (1 - u)h_3(i) + \epsilon_i, \tag{5.21}$$

where $i = 1, 2, ...32$, u is a uniform random variable on $(0, 1)$, $h_1(i) = \max\{6 - |(i - 7)|\}$, $h_2(i) = h_1(i - 8)$, $h_3(i) = h_1(i - 4)$ and $\epsilon(i)$'s are the i.i.d. standard normal variables.

Five sample waveforms from each of the three classes is shown in Figure 5.14. The critical differences between the signal classes has been previously proven [86] to be identified better using TF maps in comparison to spectral analysis tools and hence is considered a suitable candidate for robust assessment of the proposed AD-ML kernel-based analysis. In order to

evaluate the performance of the proposed scheme, a 3-class classification problem using the triangular waveform data (with signal length = 32) generated using Eqns. 5.19, 5.20 and 5.21 are considered. From the Figure 5.14 it can be verified that the signal is highly random and discrimination between the three classes is not quite obvious.

In this analysis, 1000 sample waveforms are generated from each of the signal class and system performance is assessed. The feature vectors were extracted based on the proposed algorithm and these along with the class information are fed to a linear classifier. A correlation-based analysis, for the chosen feature vectors, is performed for each class group in order to validate the independence of the features before the classification process. In addition, since the triangular dataset is generated stochastically, the simulations are iterated 20 times in order to efficiently assess the discrimination performance.

Table 1 lists the performance accuracies (defined in terms of the misclassification error) obtained using a linear discriminant analysis (LDA) classifier by means of leave-one-out (LOO) cross-validation and resubstitution approaches. It can be noted that features extracted from a kernel-mapped AD representation performs significantly better than a direct AD-based analysis for the considered dataset. The obtained metrics are closer to those reported by Saito [86]. Brieman theoretically proved that the minimum Bayes error obtained using the resubstitution method is 14% for the 3-class problem and no empirical procedure is likely to produce better results.

Investigation on the features reveals that an estimate of the relative cross-term energy distribution in AD varies significantly between different classes in comparison to auto-term concentration. Quantitatively speaking, the inclusion of cross-term feature vector provided an improvement in the overall classification accuracy (ranging between 20-30%). This is an interesting perspective in the area of cross-term (previously tagged 'unwanted') research and is obtained during our preliminary investigation on useability of frequency component interactions in a multicomponent signal. Further analysis in this field of study, can provide

Table 5.1: Misclassification Error obtained for the 3-class triangular waveform problem

Transformation Used	Classifier	Misclassification Error
Direct AD-based	LDA-LOO	32.7%
Direct AD-based	Resubstitution	27.7%
AD + Fisher Kernel	LDA-LOO	16.9%
AD + Fisher Kernel	Resubstitution	15.3%

new directions for modeling these terms and can also facilitate efficient characterization of real-life NS signals.

5.4 Chapter Summary

In this chapter, certain novel kernel modifications in AD for NS signal analysis signals were introduced. The first of the two contributions, in this chapter, addresses the issue associated with obtaining good TF resolution and reasonable TF localization in smoothed distributions. Suitable modifications on two existing smoothing kernels (Choi-Williams and Margenau-Hill) were proposed. The evaluation of the performance of the proposed kernels indicates the improved resolution characteristics.

The second contribution is in the area of TF learning machines. These classes of kernels have recently gained increased interest owing to the fact that a kernel induced non-linearity is introduced in the AD decomposition stage for extraction of robust and representative feature vectors. Further, an investigation on the usability of cross-terms generated when the signal is bilinearly transformed in the TF domain.

The work till now concentrated on exploiting the richness in AD for developing schemes that facilitates efficient NS signal analyses and their characterization. AD being a 2-D domain

characterized by a time-varying autocorrelation function, the usability of this domain can be used to characterize certain real-life signals that show some inherent correlation behavior. In the next two chapters, this aspect of AD will be assessed using a novel discriminative feature space for certain biomedical examples. In order to facilitate this, two biomedical datasets that carries inherent rhythmicity: walking gait and pathological speech are used. In addition, there exists a possibility of existence of long term correlation information in these signals and hence, are used in this work for demonstrating the applicability of AD.
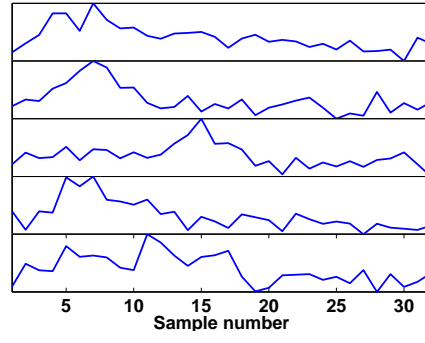
(a) Class 1
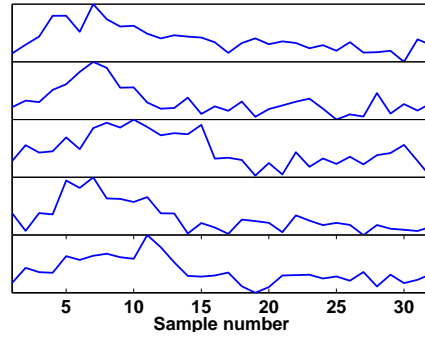
(b) Class 2
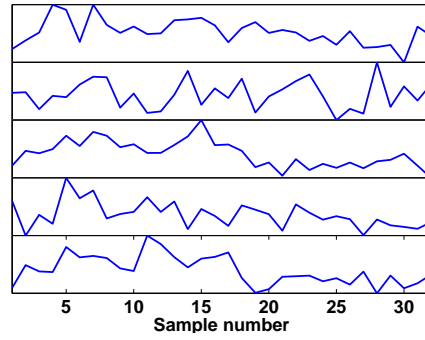
(c) Class 3

Figure 5.14: Synthetic dataset: Five sample waveforms for each class of the Triangular dataset, generated using Equations 5.19, 5.20 and 5.21 respectively, and x-axis represents the sample number.

Chapter 6

AD Analysis of Gait Pattern

The onset of a neurological disorder, such as amyotrophic lateral sclerosis (ALS), is so subtle that the symptoms are often overlooked, thereby ruling out the option of an early detection of the abnormality. ALS, also called Lou Gehrig's disease, is a neuromuscular disease characterized by progressive weakness resulting in abnormalities in dexterity and during walking. As the motor neuron system degenerates, ALS often manifests in the form of muscle weakness, muscle atrophy, fasciculation, and multiple combinations of corticospinal tract signs [87, 88]. ALS is one of the severe neurological disorders among adults and a majority of the affected individuals die within five years of diagnosis [89]. At the onset of the ALS disorder, the motor neurons begin to deteriorate thereby affecting the strides from one gait cycle to the next. Subjects with ALS also show a loss of muscle strength and coordination along with other symptoms such as difficulty in breathing and swallowing which appears in the later stages. These symptoms can often be misconstrued with other disorders. Hence, the disorder goes undiagnosed till the final stages of complications. In order to effectively pre-screen the presence of ALS, an AD-based scheme for altered gait analysis is presented in this chapter, as shown in Figure 6.1.
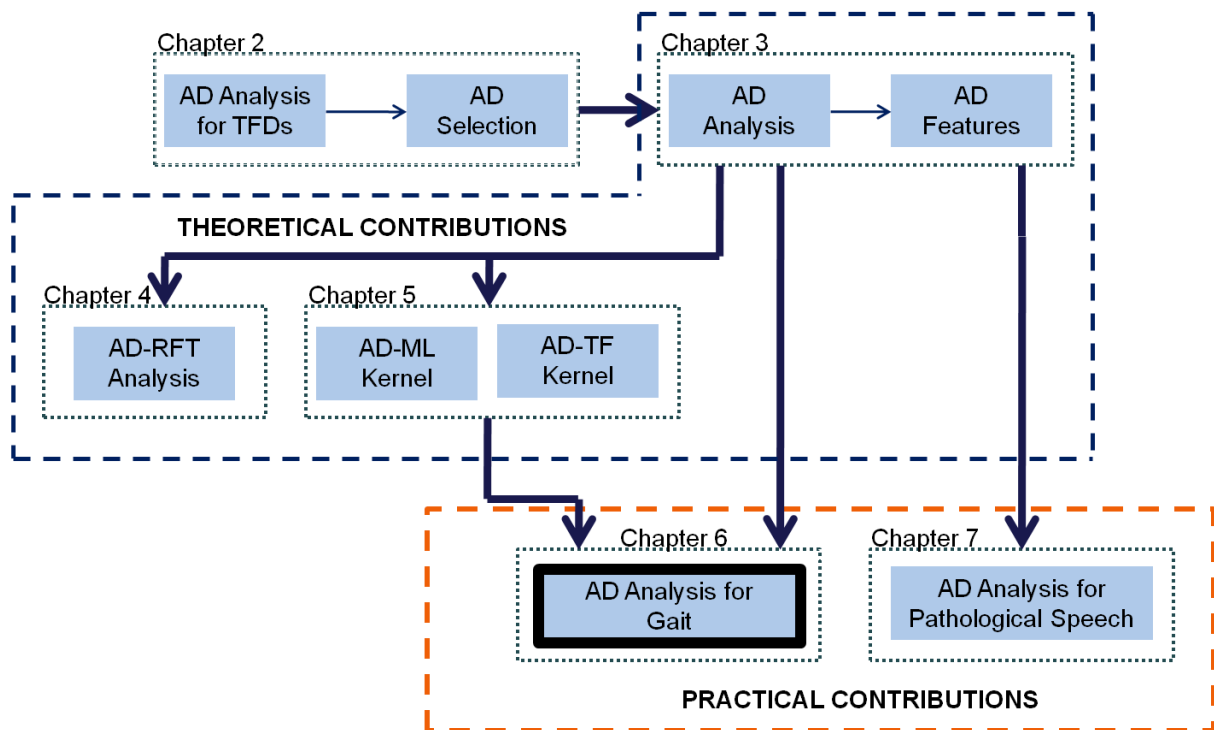
Figure 6.1: Chapter 6 - AD Analysis of Gait Pattern

6.1 Previous Studies

Existing diagnosis measures, to rule out other medical conditions, include blood tests, EMG analysis, genetic testing (to check if there exists a family history for ALS) and spinal tap (lumbar puncture). From this, it can be observed that the task of detection of this disorder is challenging and difficult. Most importantly, currently, there are no treatment measures available for ALS. The main objective of the current research work is to develop an effective screening algorithm (for ALS) that provides an objective metric for the gait data.

In general, the gait stride interval (or the gait cycle) values differs among individuals. This is usually due to evolved innate preferences between individuals [90] and these differences are usually characterized by measuring some of the gait parameters such as stride time, duration of the stance phase, swing phase and the total gait cycle. The available experimental setups [91] for gait recording is as shown in Figure 6.2. Depending on the distribution of stance and swing phases, bipedal gait can be divided into four categories (see Figure 6.3): walking, race-walk, running and sprint gait [1]. The stride time, is defined as the time between initial contact of one foot to the successive contact of the same foot. A single gait cycle or stride is usually characterized by the existence of two phases: stance and the swing phases. The interval in which the reference foot is in contact with the ground is called the stance phase and when this activity constitutes about 60% of a single cycle, the corresponding gait is referred to as the walking gait [92]. While, the swing phase contributes for the remaining gait cycle and is computed from the difference between the gait cycle and stance phase, when the foot has no contact with the ground [93]. Throughout the scope of this dissertation, the walking gait is considered for the analysis. The recorded gait pattern contains information about the fluctuations from one stride to the next.

Analysis of gait in order to identify the disease along with providing a risk assessment can provide significant preventive and diagnostic benefit, especially among the aged popu-

Sensor

**Gait
Setup**

| Floor gait plate |
| Footswitch system |
| Motion recording sensor |
| Rotation Sensor |
| EMG |
| Biomechanical Sensor |

| Hip Joint |
| Knee Joint |
| Tibia |
| Ankle Joint |
| Calcaneus/ Navicular |
| Forefoot |
| Toes |
| 3D Motion |

| Analog-to- Digital Converter |

**Gait
Parameters**

| Elapsed time |
| Stride Interval |
| Swing Interval |
| Stance Interval |
| Double Support Interval |
| Angular Measurements |

Figure 6.2: Gait Acquisition Setup
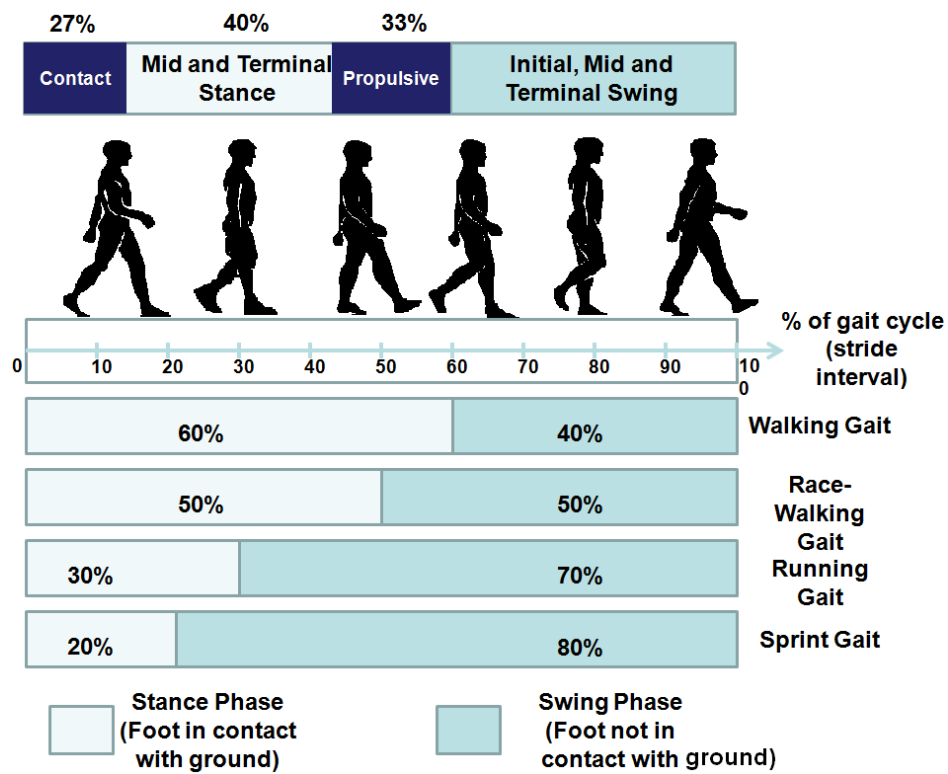
Figure 6.3: Typical Gait Cycle for Different Gait Types [1]

lation. Some of the available means for assessment are based on clinical observations and patient self-reports. These techniques provide analysis reports that are qualitative, which limits the access to the general care population. In order to facilitate automated diagnosis of various neuromuscular disorders that would assist clinicians, application of pattern analysis and machine learning has gained increased importance in recent time [94–97]. The underlying principle of operation of these pattern selection schemes are: reducing redundancy and obtaining maximally relevant features that identifies the characteristic features of a given signal.

The literature [98–101] suggest that use of signal processing techniques could aid in obtaining suitable characterization of gait signals. Also, such analysis can facilitate understanding of the motor control mechanism in human body and is useful for monitoring neurological disorders. Hausdorff et al [102–104] analyzed the variation in gait for neurological diseases. Their study revealed longer stride interval for ALS patients in comparison to the normal gait. They reported a reduction in correlation behavior between successive stride intervals, for abnormal gait signals, using the detrended fluctuation analysis technique. Certain other studies [105, 106] employed the wavelet-based fractal analysis and TF matching pursuit algorithm to analyze the gait of subjects with neurodegenerative diseases. Their research reported an increase in fractal dimensions (measured in the wavelet domain) for subjects with neurological disorders.

These studies suggest that gait characterization can contribute significantly towards detection of the presence of pathological activity (like ALS, cerebral palsy, etc) in a subject, thereby facilitating diagnosis. An early detection based on gait parameters aids in efficient treatment and can be extended for patient monitoring in long term care. There exist two kinds of approaches for gait classification: through statistical analysis and using artificial intelligence techniques. Cunado et al [107] proposed a scheme that uses the phase-weighted magnitude for the change in inclination of the legs, obtained using Hough transform, as

features for the classifier. Some of the other pattern recognition algorithms for gait analysis include neural networks (NN) [94,95,108–111], support vector machines (SVM) [96,108], and radial basis functions [112]. The NN-based schemes use a combination of the basic gait data parameters, kinetic data and kinematic data based feature vectors for classification analysis. The kinematic readings and the statistical parameters are extracted for schemes employing SVM and radial basis function classifiers.

Table 6.1 shows the list of feature vectors contained in each of the gait parameter category mentioned above. The main purposes of these studies were to obtain characterization of young and old subjects, by extracting features from their respective gait data. That is, computing features based on temporal-distance, kinematic and kinetic characteristics, and then using them in SVM- and NN-based classifiers.

The gait characteristics were also analyzed by use of certain biometric identification systems, some of which use Doppler signals, with an aim to achieve better classification performances. These schemes [113–115] extract micro-Doppler (MD) features, obtained using Doppler scattering concept, in the joint TF domain. Here, techniques such as spectrogram, linear basis decompositions were applied to the MD signals and the frequencies of rotation and the corresponding phases used as feature vectors. Also, certain other schemes employing, independent component analysis (ICA) has been used in conjunction with spectrogram to extract independent basis functions for use as features in a classifier [116]. Boulgouris et al [117] proposed a scheme for gait representation and recognition. In this scheme, the features were extracted using Radon transform of binary silhouettes and subsequently subjected to a linear discriminant analysis (LDA) for classification. Wu et al [118] assessed the gait variability in patients with Parkinson's disease using a non-parametric Parzen-window method approach and reported a maximum classification accuracy of 90.32% using the SVM classifier.

Table 6.1: List of extracted feature vectors for available gait parameter categories

Gait Parameter Category	Extracted Feature Vectors
Basic Gait Data	walking speed, stride length, stance time, double-stance time
Kinetic Data	variety of foot-ground reaction force measurements, mediolateral and fore-aft shear forces
Kinematic Data	knee and ankle joint angles in frontal, transverse and sagittal plane, 3-D motion analysis readings, joint range of motion for planar movement
Statistical Data	centroid, mean-square abscissa, variance of abscissa

6.2   Motivation

Irrespective of the existence of many schemes, very few schemes that target characterization of gait data are available. In terms of classification applications, there still exists a need for robust feature extraction techniques and efficient discrimination algorithms for analysis of such time-varying signals. Further analysis on gait can assist in quantifying the non-stationarity in gait that can in turn facilitate discrimination between normal and abnormal gait, for applications related to diagnosis of neurological disorders. Studies also indicate that the stride interval variability exhibits a complex, non-linear behavior [119], because of the non-linear dynamics of the human system. For analyzing signals with time-varying spectral characteristics, the time- or frequency-based methods needs to be used in conjunction with suitable segmentation algorithms (for fixed length or variable length segments) in order to identify the TF connection in the signal. Hence, in order to retain the TF relation over the entire duration of the signal, without the need to suitably segment, TF techniques are often preferred. For this purpose, the AD is investigated and exploited for obtaining a novel

discriminative analysis scheme for gait.

The main motivation of this work is to propose a robust characterization scheme for the time-varying gait. For this purpose, the representative space of the AD is exploited, and signal representative feature vectors are deduced from the RDRs. Here, the stride and swing intervals of a gait signal obtained from individuals are analyzed for automating discrimination among healthy and ALS-affected individuals.

In addition to changes in fluctuations in stride intervals among individuals, there can exist differences in gait cycle owing to gait asymmetry [120] owing to presence of disorders and aging. However, in order to make the proposed work less dependent on the asymmetry behavior, the current work aims at characterizing the correlative behavior of the fluctuations in stride for a normalized foot signal. This step ensures that although there can exist an asymmetry among different individuals (for signals obtained from right and left foot), the overall inherent correlation is least biased. Hence, asymmetry is not discussed during analysis of walking gait data during the scope of this work.

6.3 AD-Based Characterization of Gait

The gait stride interval variability poses difficulty during analysis among different disorders due to the inherent random behavior. In order to examine the dynamics of the signal, the gait time series data is first suitably transformed into a higher dimension where certain signal-specific feature sets, based on relative energy concentration among the signal and interfering terms, can be defined.

Significance of AF properties in Gait analysis

The AF is the dual of the WVD in the sense of the Fourier transform, and hence the properties of the distribution can be extended for characterization of this function as well. Two of the AF (and for the TFD) properties namely the shift invariance and marginals play a

significant role during analysis of time-varying gait signal. Since the stride-to-stride fluctuations changes in a non-linear fashion in an individual's gait cycle, the magnitude of the AF should be invariant irrespective of signal shifts in the time- and frequency- planes. Hence, there exists a need for distributions that provides robust performance characteristics for time and frequency shifts, thereby the signal components and their shifted versions look the same regardless of the specific location in the TF plane.

Also, since AF satisfies the marginal property, the temporal and spectral auto-correlations can be directly obtained by the points of intersection along the t and θ axis respectively. The marginal property aids in computation of the instantaneous features of the stochastic signal from the corresponding probability density functions (extracted in time and frequency planes) [121]. More information on the properties and the governing equations can be obtained from [14, 45]. Furthermore, it is interesting to note that in AD, the maximum signal energy concentration is about the origin and the diagonal entries contains most of the signal terms. And, depending on how close the different dominant frequency components are, the cross-terms are generated due to signal interaction (in a multicomponent signal) and are located away from the origin (represented by off-diagonal entries). This sort of a representation can be probed for investigation on cross-term distribution among normal and abnormal subjects gait data.

6.3.1   Dataset

The gait data used in this analysis is obtained from the public database archive available at Physionet. A group of patients devoid of any known neurological disease conditions, comorbidities and who were not under medications that might have an effect on the gait pattern, contribute to the control subject (CO) dataset in Physionet ("http://www.physionet.org/physiobank/datab

Data Collection Protocol

Each of the subjects (PD, HD, ALS and CO) was instructed to walk at their normal pace along a 77m long hallway for 5min (300s) without interruption. The pressure-sensitive captors were placed on each of the subjects' right shoe in order to facilitate computation of the gait. The subject generated force (applied on the floor) signal is stored using a light weight, ankle-worn recorder (dimensions: 5.5x2x9cm; weight: 0.1kg) was worn on the ankle cuff of each foot. A conductive polymer layer sensor transducer is used to convert the force signal into electrical signals. A thin (thickness < 0.05 inches) commercially available (Interlink Electronics) battery-operated sensor that offers fast dynamic response, temperature insensitivity and ease of interfacing is selected for the footswitch system [102]. The overall error in estimation of gait intervals using this sensor is validated to be within 3%.

The recorder signal output is sampled, at a rate of 300Hz, using an on-board 12-bit analog-to-digital converter (ADC). From the transducer output, the stride interval is calculated using a three step process [122]. First, the local minima, local maxima and the rising and falling edges are located. In the next step, the local minimum is estimated for each edge (rising and falling), in order to be used as a local baseline. This step ensures that the step-to-step changes in the interface do not alter the absolute minima. Finally, the time from initial contact to the next initial contact of the same foot for each gait cycle (stride interval) is defined as the time when the signal drops below a small offset value defined relative to the local minimum computed in the second step. Additional information about the gait interval computation from the sensor output can be obtained from [122].

More details about the outdoor data collection can be obtained from [103]. Table 6.2 shows the average self-chosen walking speed amongst different individuals [93].The stride to stride measurements are obtained for a total period of 300 seconds. For the case of ALS, the severity metric used defines the time since the onset of the disease. Almost 80% of the patients

displays moderate severity of the disease.

Sixteen healthy subjects (2 men and 14 women) aged 20-74 years, and thirteen ALS subjects (10 men and 3 women) aged 36-70 years contributed to the gait dataset. Literature [123] shows that ALS most commonly affects people between 40 and 60 years of age, which justifies the availability of a specific age group for ALS group. Further, as mentioned in Section 6.2, since the proposed approach aims at quantifying the inherent correlation in stride intervals, the presence/ absence of ALS can be suitably characterized while a relatively low gender-bias exist. For the subjects with ALS, a score that represents the time (elapsed number of months) since the onset of the disease is provided in the database for use as a marker to identify different ALS subjects. This score ranged between 5.5 and 54 with a mean of 21.3 months. Two of the subjects had familial ALS and the medication intake for all the ALS subjects remained unaltered. Heights and weights of the control and ALS subjects were not significantly different and the presence/absence of a medical condition (normal or abnormal) was determined by physical examination and detailed review by a qualified physician. These ALS subjects were not using a wheelchair for mobility and were not diagnosed with any other ailments that might affect the gait [97]. The subjects that contributed to the database were selected based on their abilities to walk independently and they were instructed to walk at a self-determined pace in order to reduce the walking variability for each subject considered individually.

6.3.2 Time-Frequency Mapping of Gait

For the case of gait, it is to be noted that the fluctuations measured between stride signals from the left and right foot characterizes the time-varying gait. A reasonable gait signal representation should therefore satisfy the following:

Table 6.2: Estimated self-selected walking speed based on different test conditions and amongst individuals

Walking Conditions (Based on age: 6-80 years)	Normal self-selected gait speed (metres/second)	
	Healthy Men	Healthy Women
Indoor	1.3 - 1.6	1.3 - 1.5
Outdoor	1.18 - 1.34	1.1 - 1.29
Age Related reduction	Between 0.1 % and 0.7 % per year	
Under ALS conditions	Variable (includes healthy gait speed range)	

- the ability to capture the long-term correlation information available in the gait signal while preserving the TF connection in the signal and

- provide efficient transformation into a domain where the extracted signal features can aid in discriminating between normal and abnormal gait to facilitate prognosis, along with allowing estimation of the relevant instantaneous parameters.

In order to suitably define the time-varying spectrum of the gait while characterizing the long-term correlation information available, the AD is used in representing the signal. For any given signal, $x(t)$, the AD-map is obtained by using the AF as defined in (3.8). That is, AD is utilized to exploit the unique advantages associated with TF analysis, like localization of spectral components for extraction of instantaneous frequency and group delay features. Very few researchers have utilized these properties of the AD in their analysis. Research [39,44] and our investigation of this domain helped us in deducing the following significant characteristics that holds the key for feature extraction:

1. The cross- or the interfering terms appear between the signal components form the off-diagonal entries thereby the auto-terms are positioned around the origin and cross-terms

are spread outside the signal terms.

2. Most of the signal's energy is defined about the origin and this can be verified by the presence of maximum valued coefficients that appear as the diagonal entrants in the AF.

3. The AD-mapped signal is symmetric about the origin and the structure spread is proportional to the temporal coherence of the signal terms. It is also interesting to note that the more uncorrelated (or random) the signal is, the less is corresponding spread among auto-terms.

6.3.3 Methodology

Scheme I

The force signal obtained across the sensors (one under the heel, the other under the forefoot and the toes) are combined during the gait recording and is then digitized. From the recording, the time from initial contact of one foot to subsequent contact of the same foot (or the stride time of the gait cycle) is computed by using an algorithm to locate the initial contact time and the changes in the slope of the force [102]. It has been reported that the weight applied at the sensor points varies owing to the internal dynamics of the human body. This results in the observed non-stationarity in the stride time reading.

Also, studies [124] about human locomotory behavior suggests that the fluctuations of stride-to-stride intervals, derived from continuous gait recordings, exhibit randomness and are nevertheless possessing long-range correlations. That is, gait stride interval time series data do not necessarily carry obvious deterministic signatures (in general), in contrast to their expected rhythmic pattern. As mentioned, Fourier analysis is a good technique to model rhythmic data; however, when the signal becomes NS, TF approaches (such as AD-based schemes) can provide a better characterization for the correlative behavior of the gait signal.

To facilitate characterization, the signal needs to be mapped on to a suitable representation space. The modified signal coefficients are then used for extracting relevant feature vectors from the normal and pathological gait.

In this work, the use of energy distribution across the signal and the cross-terms (additional terms generated due to interaction between different frequency components) is introduced as feature vectors. Figures 6.4 and 6.5 give the 2-D AD representation of the gait data of a ALS male and a healthy female respectively. The AD-plot of the transformed signal represents the energy distribution among the auto- and cross-terms. From Figures 6.4 and 6.5, in comparison to the normal-gait, region of maximum energy concentration (corresponding to reduced spread about both the axes) appears close to the origin in ALS gait. For example, the spread about the origin in the time-lag axis of ALS-gait plot (between -10 and +10) is less than the corresponding spread for the control gait (between -20 and +20). This implies that the auto-terms in ALS gait is concentrated more around the origin and in addition, from figures it can be verified the energy spread decreases moving away from the origin. Similar inferences were derived from analyzing different subject (gait) data for both the groups.

Scheme II

For statistical knowledge discovery, kernel-based methods have attracted many researchers, mainly because of the fact that, kernels provide a linear or a non-linear map of the signal in the higher dimension (Hilbert) space by computation of simple dot product in the input (feature) space. In this work, a transformation scheme in AD that exploits the benefits of the machine learning kernels is introduced, in order to reduce the number of computation steps needed to deduce an optimal feature space.

AD-SVM kernel:
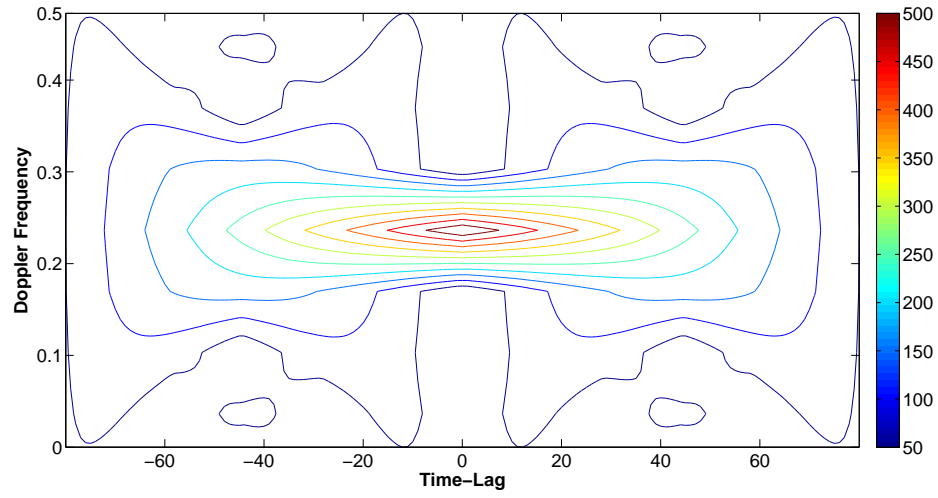
Figure 6.4: 2-D contour plot of AD representation of the gait data of a 43-year old ALS male
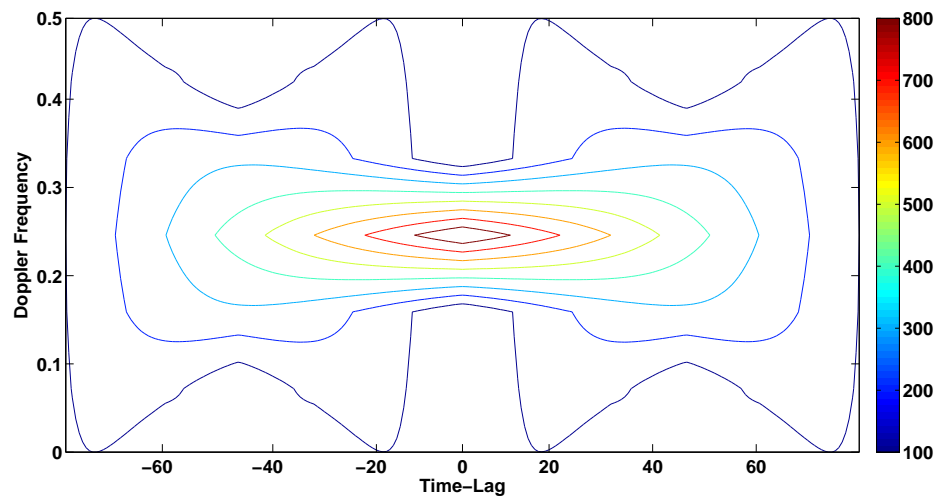
Figure 6.5: 2-D contour plot of AD representation of the gait data of a 57-year old healthy female

132

For any given input space, $X \subset R^d$ with vectors x and z, the kernel is defined as:

$$K(x, z) = \langle \phi(x), \phi(z) \rangle, \tag{6.1}$$

where ϕ is a non-linear (or a linear) map from the input space X to the feature space F, and $\langle \rangle$ is an inner product. Such a representation is also symmetric and reversible i.e. $K(x, z) = K(z, x)$ and satisfies Cauchy-Schwartz inequality:

$$K^2(x, z) \leq K(x, x)K(z, z). \tag{6.2}$$

The main idea underlying the choice of a SVM kernel, is that the derived kernel should maximize margins between hyperplanes, in the feature space, in order for to efficiently distinguish between different classes. This concept is extended to TF representation, so that a better representation of critical features is achieved. There exists three main types of SVM kernels that facilitates maximization of margins between hyperplanes: radial basis function (RBF), polynomial kernel and the sigmoid kernel. Of these, the all-subset or the polynomial kernel is chosen in this work since it offers reasonable separation between classes of data whose intra-class relation is non-linear. A polynomial kernel is represented by the following equation,

$$K(x, z) = (\alpha + x \cdot z)^d. \tag{6.3}$$

Here, $x \cdot z$ represents the dot product between x and z, α is any scalar quantity and d denotes the order of the kernel.

Selection of parameters: The choice of 'α' and 'd' parameters are governed by AD-map of the signal. To define these values, the AD-map is divided into bands based on relative energy distribution ratios. This ratio aids in the relative characterization of the energy distributed about the origin and hence, provides the relative energy between auto-

and cross-term distributions from the AD-map. From this representation, the region of auto- and cross-term overlap is approximately identified using the energy spread factor and the maximum energy distributed between successive bands. That is, in AD the energy map usually resembles a bell-shaped distribution and in regions where there are more cross-terms, the maximum energy decreases alongside a decrease in the signal spread. Once the region (band number 'n' from the origin) is obtained, the maximum energy for all signals under consideration is computed. Thereafter the median value obtained from the ranked maximum energies, is then assigned to 'α' and the parameter 'd' is computed as 2^n.

The SVM kernel function is then incorporated in the characteristics function (Eqn. 3.1), thereby defining a new TFD for NS signal analysis,

$$M_{SVM}(\theta, t) = \int K \left\{ x \left(\theta + \frac{t}{2} \right) x^* \left(\theta - \frac{t}{2} \right) \right\} d\theta \, dt. \tag{6.4}$$

The input signal is mapped on to the Hilbert space by using the proposed AD-SVM kernel function. From the transformed signal, the newly derived feature vector set is extracted.

Feature Space

Prior to defining the feature plane, the AD-map is divided into 'n' equal bands and the relative energy distribution is quantified using the proposed relative distribution ratio (RDR) for each segment (i = segment number) given as,

$$RDR_i(n) = \frac{1}{nT_E} \left\{ \sum_{k=1}^{n} [A_k(\theta, \tau) - A_i(\theta, \tau)] \right\}. \tag{6.5}$$

Here, $A_n(\theta, \tau)$ represents the total energy in the nth band i.e. total energy for all values of θ and τ in the nth energy band, and T_E represents the energy in the innermost band about

the origin and is defined as,

$$T_E = \sum_{\theta_1} \sum_{\tau_1} |A_i(\theta_1, \tau_1)|^2 \,, \tag{6.6}$$

where θ_1 and τ_1 takes the following values,

$$\theta_1 = \left| \left\lfloor \frac{r}{2} \right\rfloor \pm k_1 \right| \text{ and } \tau_1 = \left| \left\lfloor \frac{c}{2} \right\rfloor \pm k_2 \right|. \tag{6.7}$$

From the RDRs, the relative energy of the two bands with reference to the region of predominant gait signal concentration (auto-terms distributed about the origin) in the ambiguity plot is computed. That is, in simple terms, the ratio represents the progressive energy distribution between the inner band and the successive outer bands. From the feature space defined using RDRs, the following 3 feature vectors were extracted:

a) maximum energy across the auto-terms,

$$f_1 = T = \sum_{\theta_1} \sum_{\tau_1} |A_i(\theta_1, \tau_1)|^2 \,, \tag{6.8}$$

b) mean of the RDRs obtained over the entire duration of the signal,

$$f_2 = \frac{1}{n} \sum_{z=1}^{n} RDR(z) \text{ and} \tag{6.9}$$

c) variance of the RDRs for each signal,

$$f_3 = \frac{1}{n} \sum_{z=1}^{n} (RDR(z) - f_2)^2. \tag{6.10}$$

The extracted feature set, being obtained from the correlation domain (defined by the AF), provides a representative of the coherent (signifying TF shifts between consecutive signal components) and the non-coherent (indicative of TF correlations) signal components in the gait. These features are then used in the class categorization, into normal and abnormal

classes.

ALGORITHM

The proposed scheme for gait discrimination among healthy and ALS subjects can be briefly characterized by the following five steps.

- **Preprocessing:** Based on the data collection protocol followed by Hausdorff et al [102] and their investigation reports, the initial 20-second data has been identified to be influenced by start-up effects and has not been included during their analysis. In this research work, the analysis in a similar fashion, in an effort to keep the influence of start-up errors on processing to a minimal level.

- **Higher Dimension Mapping:** The preprocessed signal is then normalized and then mapped on to the AD, using direct AF (Eqn. 3.8), AD-RFT (Eqn. 4.15) and machine learning kernel modified AF (Eqn. 6.3.3), and transformed multivariate signals are obtained. The advantage of using such a representation is that this domain separates out the cross-terms as off-diagonal entries facilitating discrimination among auto- and cross-terms. The autocorrelation function for the signal is computed during this mapping. The maximum energy component is obtained at the center of the matrix.

- **Extraction of feature vectors:** The transformed signal matrix is then divided into regions on the AD space based on their relative energy composition and their proximity to the origin. The number of energy bands is determined based on the length of the signal. Then the relative energy ratios (RDR) between the auto- and cross-terms are computed. These ratios, along with the total energy and the variance across the auto-terms are extracted from the AD and are used to define the feature space.

- **Classification:** Based on the feature space, the normal and ALS gait classes are defined. The feature vectors along with the class definition together are fed into linear and non-linear classifiers for performance assessment of the proposed gait characterization scheme.

- **Evaluation:** A three-stage evaluation process is carried out in this study. First, the direct AD-based feature vector set is assessed by obtaining a scatter plot of normal versus the abnormal groups. This is done to verify the validity of our approach, and to visually identify the discrimination characteristics of the deduced feature vectors with reference to the healthy and the ALS subjects. Following this, a similar assessment is carried (using LDA) for evaluating the performance of the machine learning kernels for gait discrimination.

  In the third stage of performance assessment, the classification accuracies obtained from LDA and feed-forward back propagation NN-based classifiers are analyzed and a cross-validation to verify the correctness of the obtained measures is carried out. That is, following the classification routine, the class labels for correctly classified and misclassified data, are manually cross-checked with those obtained from the database. Additional information about the classifiers [5, 125] is provided in the following subsection.

6.3.4   Classifiers

In this work, the effectiveness of the proposed feature space, for discrimination between the normal and the pathological classes, is assessed using linear (LDA) and non-linear (NN) classifiers.

Linear Discriminant Analysis

For a given set of feature vectors, LDA [5] works at obtaining maximum discrimination between classes, by using certain optimization techniques. This is achieved by maximizing the ratio of between-class variance to the within-class variance for any given dataset, thereby obtaining a maximum overall separability. The resultant decision boundaries for each of the classes is linear, thereby, grouping the different class clusters in a given subspace. The algorithm is given as follows,

Step 1: For a two-class problem, the features from both the datasets are first represented as matrices. For example, considering two features vectors for each dataset, the corresponding matrices are of the form,

$$
class1 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ .. & .. \\ .. & .. \\ x_{m1} & x_{m2} \end{bmatrix} \quad class2 = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ .. & .. \\ .. & .. \\ y_{n1} & y_{n2} \end{bmatrix}. \tag{6.11}
$$

Step 2: Following the matrix representation, the individual and group means are computed by considering the data independently and after merging them together respectively.

Step 3: In the next step, the class separability criteria is formulated by calculating the expected covariances of each of the classes, thereby the scatter coefficients for within-class and between-class cases are obtained. That is, the scatter measures are given as,

$$
S_w = \sum_j p_j \times (cov_j), \tag{6.12}
$$

138

where, cov_j is the covariance of class j respectively. For a two-class problem with S_w and S_b as the within-class and between-class scatter measures,

$$S_w = 0.5 \times cov_1 + 0.5 \times cov_2, \tag{6.13}$$

$$S_b = \sum_j (\mu_j - \mu_3) \times (\mu_j - \mu_3)^T. \tag{6.14}$$

Here, μ_3 is the mean of the global mean and is computed from the merged dataset and μ_j represents the mean of class j respectively.

Step 4: The obtained scatter measures are then optimized using the maximization of within-class to between-class covariance criteria and by computing the Euclidean distances for each test data point i.e. the test vector is classified to class j when the corresponding Euclidean distance is the smallest.

In this work, the training and the test dataset are defined using the least biased leave-one-out cross validation technique. Here, a single observation selected at random is used for validation while the remaining data is used for model training. The testing and training process is repeated until each of the available data is validated at least once. The overall classification accuracy is computed as the average of accuracies from all iterations.

Neural Networks

Artificial neural networks [125] is one of the most popular non-linear models for approximating a function of multiple inputs (features) and outputs (classes) using different neurons (programming constructs). The basic architecture of a NN-based model consists of three types of layers: input, hidden and the output. The current work uses a feed-forward NN model to classify between the normal and the ALS subjects.

Each neuron (similar to its biological counterpart) reacts based on the input stimuli. The output across a neuron is obtained by computing the weighted sum of the inputs to the

neuron, followed by adding the bias to the sum and finally feeding the sum as input to the corresponding activation function of the neuron. Mathematically,

$$Output = A_c \left(\sum_k w_k \, I_k \; + \; bias \right),$$ (6.15)

where A_c is the activation function of the neuron, w_k is the weight of the kth in-edge, I_k is the input carried across the k_{th} in-edge, and bias is the bias of the neuron.

For a given set of training dataset (input-output pair), the weights and bias are evaluated iteratively until a desired minimum of sum of squares error is reached. During the training phase, the weights and the bias were computed for all training datasets, and a simple optimization is carried out using the back-propagation (or the gradient descent) algorithm. In NN-based classification, the given data is divided into three subsets: training, validation and test set. The training and the validation subsets are used for optimizing the model parameters i.e. iteratively adjust the weights and bias till the desired gradient is reached. Finally, the test data is used for comparing different model performances and to finalize the model for classification. The obtained performance metrics are discussed in detail in Section 6.4.1.

6.4   Quantitative Assessment

The extracted feature vectors were used to discriminate the gait into control and ALS classes with the aid of an automatic classifier. Among the available classification methods, the LDA and NN classifiers were used for testing, to verify the validity of the proposed scheme and to assess the discrimination ability of the extracted features.

Based on the data collection protocol followed by [102], in order to minimize the start-up effects during gait measurement, the initial 20 second segment of the gait recording (for both control/healthy and ALS/pathological subjects) is removed before further analysis. The pre-
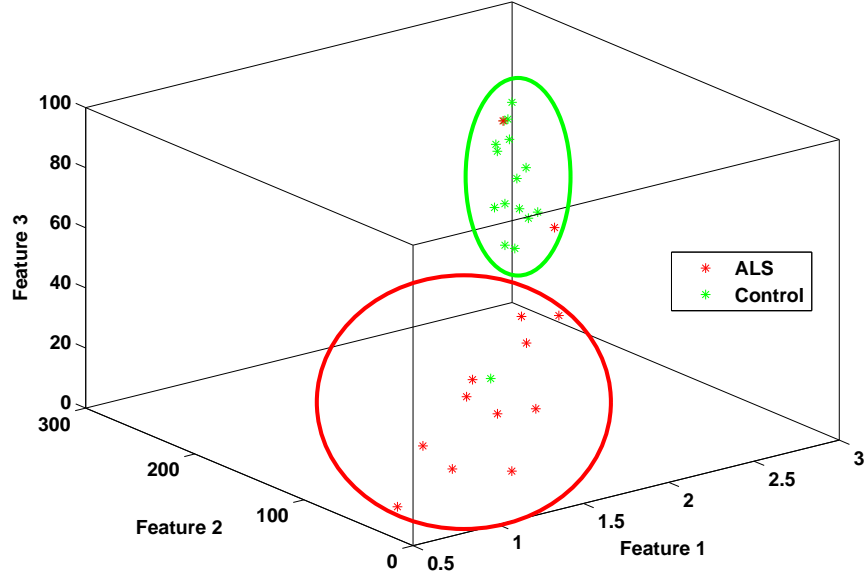
Figure 6.6: Scatter plot of extracted feature vectors: maximum energy across the auto-terms (Feature 1: f_1),mean (Feature 2: f_2) and variance (Feature 3: f_3) of the RDR (Green: Normal and Red: ALS Gait).

processed signal is transformed on to the AD and the feature vectors are extracted. In order to assess the classification performance of the proposed methodology, the feature vectors were extracted for the gait database. The feature space consists of: the maximum energy across the auto-terms (f_1), the mean (f_2) and variance (f_3) of the RDR obtained over the entire duration of the signal.

Figure 6.6 shows the scatter plot of the feature vector space obtained using direct AD analysis. By visual inspection, it is clear that using the proposed feature space offers clear separation between the two classes. This indicates the derived feature vectors suitably represent and provide discrimination among the normal and pathological (ALS) gait signals. The extracted features are then fed to NN and LDA classifiers.

6.4.1 Statistics

(A) Performance Assessment using LDA:

The classification accuracies obtained using the LDA is tabulated in Table 6.3. A leave-one-out (LOO) [126] cross-validation approach is used. Here, during the cross-validation, each case is classified by the functions derived from all cases excluding the test case. An overall classification accuracy of 89.2% is obtained for both originally grouped and LOO cross-validation approaches. The use of a kernel function, resulted in an overall improvement of about 3% (tabulated in Table 6.4). Based on the results, it is evident that the proposed sets of features characterize the gait pattern reasonably well under all test cases. These values are the current maximum that can be achieved using any gait rhythm classification. It is also interesting to note that, initially the experimentation to analyze the performance was carried out using only the f_1 and the same was fed to a linear classifier. Such a feature space provided 100% discrimination in identifying the control subject and a maximum accuracy of 40% for the case of ALS subjects. Based on the obtained numbers, it can be verified that the proposed scheme separates out the normal pattern with a higher success rate in comparison to the abnormal subjects.

The proposed scheme offers a maximum overall sensitivity and specificity of 91.7% and 88.24% respectively. Overall, only 1 healthy and 2 ALS subject data were misclassified into a different class. The computed values of false positives (Type-I error) and false negatives (Type-II error) indicate that the AD-based feature vectors offers a reasonable discrimination of the control gait from the pathological gait, for the considered set of 29 subjects.

The ALS gait data available in Physionet can be broadly categorized into two groups based on the time from the onset of the disease: mild lower extremity functional impairment (gait speed $\geq$ 1.05 m/s) and the advanced (significantly more prone to) functional impairment group. In order to obtain an extensive validation of the proposed approach, discrimination

accuracies were computed for classification between the control and the ALS classes, for the two severity groups considered separately. Overall, 1 healthy and 1 ALS subject data were misclassified in each of the group-based analysis and the obtained results were similar to those obtained from non-group based classification.

In addition to the analysis of gait signals using direct AD approach, a validation of the proposed feature extraction strategy is also performed using the AD-RFT approach (as defined in Chapter 4). Here, in order to facilitate classification, the sum of diagonal entries and the maximum auto-term energy are extracted from the AD-RFT matrix for defining the feature space. The obtained features along with the class labels are then fed to a linear classifier and validated. A maximum overall accuracy of 82.8 % is obtained with 79% specificity and 90% sensitivity. The obtained objective metrics, show the potential of the proposed sparser AD approach for discrimination between normal and ALS subjects' gait data.

(B) Performance Assessment using NN-based classifier:

The extracted features using the stride and swing intervals of normal and ALS gait were validated using a NN classifier. Here, a feed forward network is created with 10 neurons in a single hidden layer. The input and target samples were automatically divided into training, validation and test sets in this method. An overall cross-validation accuracy of 96.15% is obtained using NNs (at 3 epochs during the cross-validation process). During the iterative training of a NN, an epoch refers to a single pass through the entire training set followed by testing of the verification set. When the stopping criteria used in the training phase is minimum squared error, epoch corresponds to the maximum limit on the iterations.

From the obtained performance measure, it can be noted that the model converges fast (at the 3rd iteration) for the proposed feature vectors. A 100% specificity (and 92.3% sensitivity) is obtained for both stride and swing pattern characterization of the normal and ALS gait. These obtained quantitative metrics further validate the improved discrimination

abilities of the extracted feature vectors in the proposed scheme. Results obtained from discrimination between the control dataset from the two ALS groups did not change the performance metrics and a 100% specificity is obtained.

6.4.2 Discussion

Wu et al [97] conducted a series of analysis, for classification among control and ALS-affected subjects using an individual's gait and have reported a maximum overall accuracy rate of 82.76% and 90.32% obtained using a LDA-based and non-linear classifier respectively [97]. As demonstrated in Table 6.3, using our direct AD (or kernel modified AD) schemes a maximum overall classification accuracy of 89.2% (or 93.1%) is achieved. Figure 6.7 shows the obtained box plot for the extracted feature vector for control and ALS subjects respectively. It can be noted that the ALS inter-quartile range (IQR) falls in the mild outlier zone of the control subjects' feature plot. However, the median values and the IQR for both the data set are placed far apart. This in turn implies, the extracted feature set can be effectively used to discriminate between both the classes ('ALS' and 'Control').

In comparison to the AD-based approach, the AD-RFT method results in a relatively lower classification performance (defined in terms of the overall accuracy). In the AD-RFT approach, the conducted investigation shows that there exists a reduction in the amount of cross-terms and this could in turn be contributing to the reduced discrimination accuracy. That is, the presence of interfering terms in the AD domain could possibly contribute towards obtaining better classification between normal and ALS subjects. Further investigation can facilitate better understanding of the discrimination capabilities of this approach for NS signal analysis.

In addition, the proposed feature vectors aid in 96.15% discrimination among normal and ALS gait signals when used in integration with a NN classifier. Based on the results obtained (for ALS disorder characterization), using the linear and NN-based classifiers, it can
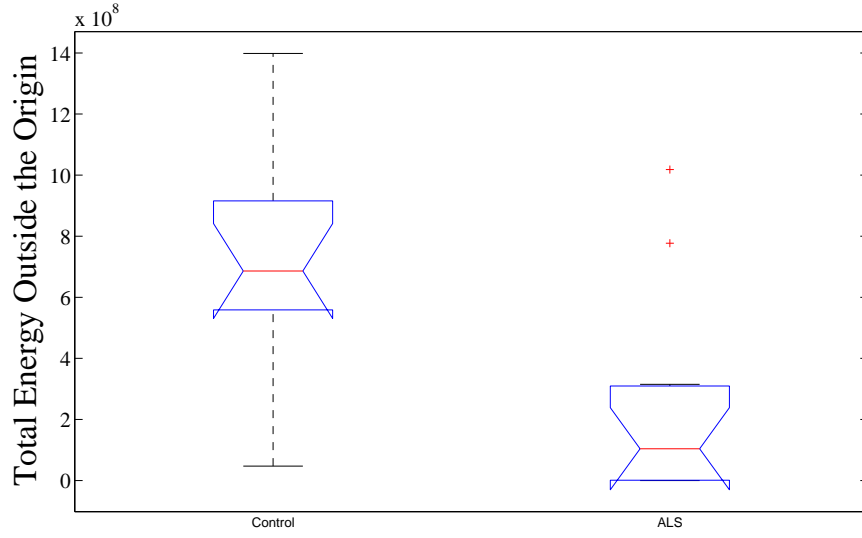
Figure 6.7: Box plot of the extracted feature vectors for the control (left) and the ALS (right) subject showing the 25th and 75th percentiles.

be verified that an overall improvement in classification is achieved in comparison to those obtained by Wu et al [97]. The obtained quantitative results were further compared with the available works on ALS classification [97, 127, 128] and are shown in Table 6.5. All of the reference works presented in the table use the normal and ALS gait data from the Physionet link.

Results obtained during quantitative assessment reveal that the obtained accuracies does not depend much on the ALS group under analysis. Also, a higher accuracy for the entire ALS group indicates the robustness the proposed scheme offers in characterizing the pathological group for different severity conditions. This is an interesting validation of the claim that the AD-based technique can provide efficient preliminary diagnosis of the disease.

Table 6.3: Classification Accuracies Obtained Using the Direct AD-Scheme

Type	Healthy	ALS	Total
Healthy	15(93.8%)	1(6.2%)	16(100%)
ALS	2(15.4%)	11(84.6%)	13(100%)

Table 6.4: Classification Accuracies Obtained Using the Machine Learning Kernel Based Scheme

Type	Healthy	ALS	Total
Healthy	15(93.8%)	1(6.2%)	16(100%)
ALS	1(7.1%)	12(92.3%)	13(100%)

From the results, it can be verified that the AD-based analysis offers new insight to suitably represent a time-varying signal in the joint TF space. Also, it is evident that the proposed feature extraction succeeds in discriminating the normal from the pathological region in a gait signal with minimal manual intervention (for training the classifier). The obtained system performance characteristics indicate an improvement in the number of correct classifications, which further validates the improved discrimination abilities of the proposed scheme. Further, it is interesting to note that the auto-term energy along with either mean or variance factor of the RDR provides similar classification performance and can hence be used invariably.

6.5 Chapter Summary

Identification of features that define the gait signal well can assist in reducing the amount of data needed for early diagnosis of neurological disorders. This study presented a robust

Table 6.5: Comparison of Classification Performance of available schemes for ALS classification

Reference	System Input	Features	Classifier	Overall Accuracy
Wu (2009)	Gait	Swing-interval turns count	LDA (or Non-Linear)	82.76% (or 90.32%)
Masood (2010)	Gait	Combination of many statistical features	Quadratic Bayes normal classifier	86.9%
Wu (2011)	Gait	Probability Density function + non-parametric parzen window approach	LDA (or least-squares SVM)	82.8% (or 93.1%)
Proposed (Direct-AD)	**Gait**	**AD-based Relative energy ratios**	**LDA (or NN)**	**89.2% (or 96.15%)**
Proposed (AD-RFT)	**Gait**	**AD-based Relative energy ratios**	**LDA**	**82.8%**
Proposed (AD-SVM kernel)	**Gait**	**AD-based Relative energy ratios**	**LDA**	**93.1%**

AD-based gait characterization scheme that incorporates a novel feature extraction algorithm and uses a standard classification scheme for discriminating the ALS- from normal-gait data. Overall, maximum cross-validation accuracies of 89.2% (or 93.1% using kernel modified AD) and 96.15% were obtained, using the linear (LDA) and non-linear (NN) classifiers respectively, for the neurological dataset available at the Physionet database. The results obtained further confirms the higher discrimination capability of the deduced feature vector set (from the proposed AD-based scheme) in comparison to the available research works on gait signal discrimination. Also, since AD is a correlation domain, the relative similarity/variability between gait samples at different instants of time can be used to facilitate a better understanding about the signal frequency interactions, thereby the corresponding cross-terms. Although the computational complexity of the proposed approach is similar or in certain cases more than the time- and frequency-based feature extraction schemes, this AD-based method shows great potential for analysis and characterization of time-varying signals such as gait stride interval. Here, the computational efficiency is defined in terms of the time involved in feature extraction from the AD. This approach demonstrates the potential of the AD for obtaining higher classification accuracy for ALS gait diagnosis. The proposed scheme has many advantages: early identification of ALS gait (and in turn prolonged life with early start of treatment procedures), minimization of overall cost (for testing and treatment) and for monitoring the progress of treatment outcomes. Although, since this is one of the preliminary studies that employs AD-space for gait signal characterization, the deduced features vectors needs further optimization for developing robust schemes for discrimination of other neurological disorders from their corresponding gait. Currently, limited information about the significance of cross-terms is obtained during the scope of this study and investigation on longer dataset can facilitate more efficient analysis on cross-terms and also for optimization of feature space. In the next chapter, another real-life signal is analyzed using the proposed AD-based scheme.

Chapter 7

AD Analysis of Pathological Speech

7.1   Motivation

Human brain possesses a natural trait to recognize voice qualities with reference to presence of abnormality, speaker's identity and identification of characteristic language patterns (including letters and words). Research shows the existence of a variety of techniques that aims at obtaining automatic speech analysis for specific applications. The presence or absence of dysphonia (pathological voice), a voice disorder resulting from impairment in the vocal organs, can be characterized using signal analysis tools. Dysphonia is a type of phonation disorder and the resultant larynx voice pattern contains traits that discriminates between the normal and the abnormal voice. Existing clinical diagnosis procedures are subjective and involve medical technicians intervention. The assessment is heavily dependant on the experience of the medical specialist and misclassifications probability exists i.e. discrepancies resulting from evaluation by different specialists and fatigue-related errors in diagnosis are probable.

In this chapter (Figure 7.1), a scheme that facilitates automatic pathological voice characterization, for discriminating between voice signal deformities and normal signal is proposed.
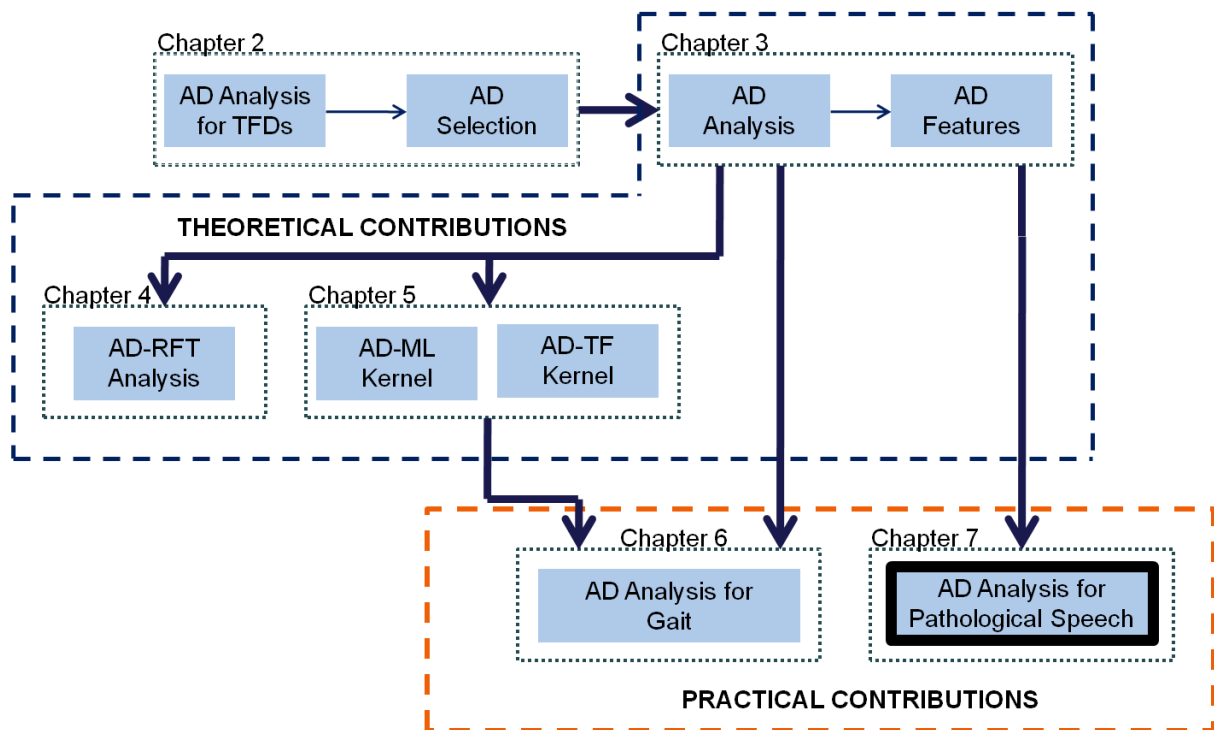
Figure 7.1: Chapter 7 - AD Analysis of Pathological Speech

7.1.1 Background

Bifurcations correspond to the qualitative fluctuations in the vocal (vibrating) system owing to parametric changes with reference to lung pressure, asymmetry between vocal folds, etc., The voice signals can be categorized into three types based on the nature of bifurcations as [129]:

1. Type I: nearly-periodic signals; the dominant frequency energy is much higher in magnitude compared to the sub-harmonics.

2. Type II: signals with more than one fundamental frequency term; their modulating (or sub-harmonics) frequency energy approach the dominant term energy component.

3. Type III: aperiodic signals.

Investigation shows that normal signals can fall into Type I or II categories, while presence of pathology, can lead to all three types of sounds. This is because, such abnormalities break down the periodicity during voice generation, making their categorization relatively challenging in comparison to discrimination among Type I and Type II signals.

Previous studies on time-domain schemes [130–132] computed temporal statistics from the signal such as jitter, shimmer, pitch variation, harmonics-to-noise ratio and residuals analysis from inverse filtering of speech signals, to distinguish between normal and pathological sounds. In the case of frequency-based analysis methods, the spectral variations over given time instants were quantified for automatic detection of pathological speech. However, there exists certain critical limitations with all these schemes, in terms of robustness, consistency of results and overall complexity. Following this, since the presence of dysphonia can induce certain NS traits in speech signals, certain joint TF approaches [35, 133] were proposed.

The importance of voice pathology has motivated a variety of studies to facilitate design of automated pathological speech detection system. Here, an automatic pathology discrimi-

nation scheme that exploits the collective advantages of the AD and employs energy based feature vectors has been proposed and is explained in the following subsections.

7.2   Automatic Speech Pathology Discrimination

Figure 7.2 shows the block diagram for an AD-based pathological speech detection scheme. To better capture the short-time characteristics, the input speech signal is first segmented before further processing. The obtained segments are normalized in amplitude and transformed into the AD-plane, using equation 3.8. A soft thresholding is then performed on the transformed multivariate signal coefficients to reduce the amount of insignificant interfering signals i.e. reduces the effect of spurious noise signals. Since the AF separates out the cross-terms as off-diagonal entries, the AD-mapped signal as such can also be used for de-noising and source separation applications, where the cross-terms are filtered out using a low pass filter (suitable kernel function).
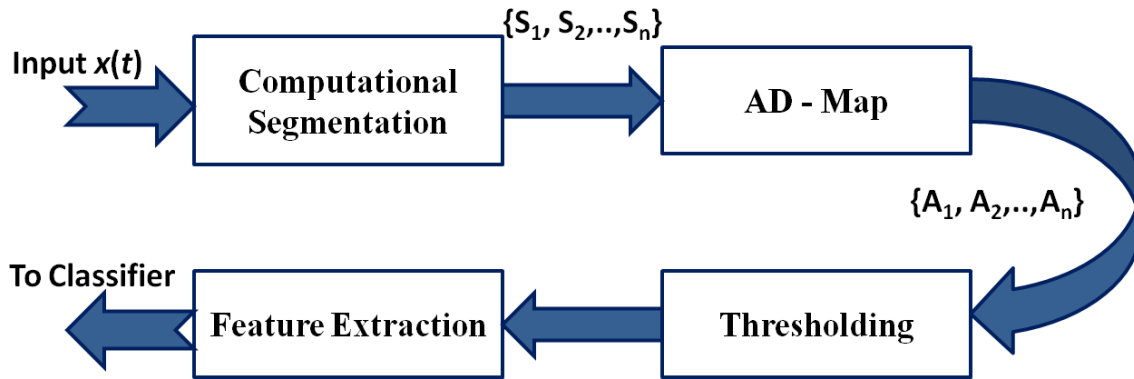
Figure 7.2: Block diagram for an AD-based Pathological Speech Detection

7.2.1 Feature Space

The relative energy of the two bands with reference to the region of predominant signal energy concentration (auto-terms distributed about the origin) in the AD plot is computed using the proposed RDRs (Eqn. 3.9). The RDRs represent the progressive energy distribution between the inner band and the successive outer bands. From the feature space defined using RDRs, the following 3 feature vectors were extracted for pathological speech detection purposes:

a) **maximum energy across the auto-terms**,

$$f_1 = T = \sum_{\theta_1} \sum_{\tau_1} |A_i(\theta_1, \tau_1)|^2 \,, \tag{7.1}$$

b) **mean of the RDRs obtained over the entire duration of the signal**,

$$f_2 = \frac{1}{n} \sum_{z=1}^{n} RDR(z) \text{ and} \tag{7.2}$$

c) **variance of the RDRs for each signal**,

$$f_3 = \frac{1}{n} \sum_{z=1}^{n} (RDR(z) - f_2)^2. \tag{7.3}$$

The extracted feature set, being obtained from the correlation domain (defined by the AF), provides a representative of the coherent (signifying TF shifts between consecutive signal components) and the non-coherent (indicative of TF correlations) signal components.

From the feature space, a novel feature set, given by the maximum, average and the variance of the relative energies across these bands (RDR_i's) is extracted. These features were then used in the class categorization (into normal and abnormal), and fed to a simple linear discriminant analysis (LDA) classifier.

7.2.2 Dataset

The data used in this study is obtained from the Masschusetts Eye and Ear Infirmary (MEEI) voice disorders database [8]. The database contains data obtained from 51 normal and 161 pathological speakers, with a variety of neurological, traumatic and psychogeni-related voice disorders.

The classification performance of the proposed methodology is validated using the MEEI datasets. The speech signals are segmented, transformed in the AD-domain and a novel set of energy-based feature vectors were extracted. The extracted features along with the class labels are classified using the LDA classifier.

7.2.3 Performance Assessment

A total of 10 unique features derived from the i energy bands in the AD-transformed signal: total, maximum, average and variance of auto-term energy about the origin, ratio of relative cross-term to auto-term energy, and 5 other features obtained as ratios of outer energy bands with reference to the auto-term energy. Figure 7.3 shows the classification performance for normal and abnormal speakers based on the number of feature vectors. Based on their discrimination capability, the best set of feature vectors (variance, ratio of cross-term to auto-term energy and total auto-term energy) were then selected using a step-wise iterative search.

Table 7.1 shows the classification accuracies obtained using the proposed AD-based pathological speech detection approach. From the Table, it can be observed that out of 51 normal signals, 48 were classified as normal and 157 out of 161 abnormal signals were correctly classified. Overall, only 3 normal and 4 abnormal data is classified incorrectly into a class and also an overall cross-validation classification accuracy of 96.8% is obtained.

AD being a symmetric intermittent TF domain, the number of computations is reduced to
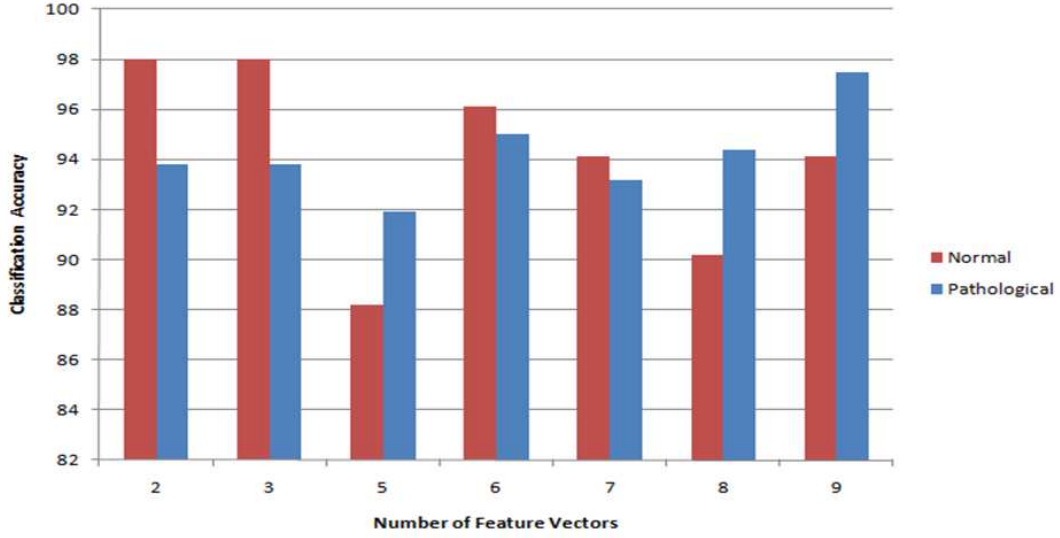
Figure 7.3: Classification Accuracy versus Number of Feature Vectors

half and hence, this approach is computationally less complex (O($N/2$ X N)). From the results, it can be verified that the proposed feature extraction scheme succeeds in discriminating the normal from the abnormal speech signal with minimal complexity, defined in terms of overall dimensionality and feature extraction process. In comparison to the existing high-performance approaches, the AD-based method performs reasonably well with energy-based feature vectors.

The earlier works on pathological voice classification, for the MEEI database, were used as pointers to validate and assess the performance of the proposed method. Previously, Umapathy et al [12] and Ghoraani et al [13] proposed TF based scheme for discrimination in pathological voice database and reported a maximum classification accuracy of 93.4% and 98.6% respectively. Table 7.2 shows the comparison of the AD-based Scheme with the state-of-the-art algorithms in terms of the dataset, methodology and classifier complexity. It can be validated that the proposed scheme offers performance comparable to or in certain cases better than the available joint TF feature extraction schemes. In addition to the direct AD-

based approach, analysis was carried out using the kernel modified AD features. An overall accuracy of 93 % is obtained, using two of the previously defined feature vectors (Section 7.2.1). However, additional investigation is needed in order to optimize this kernel learning scheme for facilitating efficient pathological speech discrimination.

Table 7.1: Classification Performance of the AD-based scheme using LDA (leave-one-out) Approach

Method	Groups	Normal	Pathological	Total
Original	Normal	48	3	51
	Pathological	4	157	161
%	Normal	**94.1**	5.9	100
	Pathological	2.5	**97.5**	100
Cross-validated	Normal	48	2	51
	Pathological	4	157	161
%	Normal	**94.1**	5.9	100
	Pathological	2.5	**97.5**	100

7.3 Chapter Summary

An AD-based approach for characterization of pathological speech signals is presented. Using the proposed methodology a maximum accuracy of 97.5% (Overall accuracy: 96.8%) is obtained for identifying the pathological voice from normal voice, for the MEEI voice disorders database is obtained. Owing to the simplicity this domain offers in understanding the auto- and cross-term distribution, further research can aid researchers in obtaining a better understanding of the cross-term distribution and their probable usefulness in classification of real-life signals. The proposed discriminative analysis can be extended to obtain robust feature spaces for optimizing classification performance in many other applications.

Table 7.2: Comparison of AD-based Scheme with State-of-the-art Algorithms

Reference	Data	Approach	Classifier Type	Accuracy
Parsa (2001)	Patho Database (53 N and 175 P)	LP Modeling	LDA	68%
Godino (2004)	MEEI Voice and Speech (53 N and 82 P)	Cepstral Parameters	LVQ	96%
Umapathy (2005)	MEEI Voice (51 N and 161 P)	TFD as feature vectors	LDA	93.4%
Ghoraani (2009)	MEEI Voice (51 N and 161 P)	TFD + NMF	k-means Clustering	98.6%
AD-Based (Proposed)	**MEEI Voice (51 N and 161 P)**	**Relative Energy Distribution Ratios**	**LDA**	**96.8%**

Chapter 8

Conclusion and Future Work

8.1   Overview

Practical signals are inherently NS and non-deterministic in nature. Owing to the non-availability of any apriori information, the challenges in their analysis tend to be multi-fold. The main objective of this dissertation has been to provide novel signal processing schemes that can suitably address the shortcomings of the existing approaches for time-varying signal analysis.

8.2   Theoretical Contributions

Signal Analysis in AD:
For time-varying signals, the time- or frequency-based methods fail to characterize the underlying characteristics. Hence, in order to retain the TF relation over the entire duration of the signal, the intermediate TF representations using AD, were investigated and exploited for obtaining a novel discriminative analysis scheme. In addition, a novel set of energy-based feature vectors that depend on the characteristic function of a TF representation were proposed. Further, an investigation on the usability of cross-terms has been provided.

Although, from a visualization perspective, the presence of cross-terms often plagues the representation. However, our investigation proves that for characterizing time-varying signals, the information available in the cross-terms can be suitably exploited. In addition, since the apriori information for most real-life signals is not available, a suitable quantification of these terms can provide one additional dimension to define the signal model.

AD Transformation using Ramanujan Sums:

Research in signal processing shows the availability of a variety of transforms for suitable representation of the given data in the higher dimensional space. The underlying difference in all these techniques is in their basis functions. Here, a novel class of RFT based TF functions, based on concepts of number theory, is introduced. Analysis shows that this new class of transforms offers very high immunity to noise interference. Further, a 2-D formulation of the RFT function has been proposed. Simulations using synthetic signals show the potential of this approach for obtaining relatively sparse TF-equivalent representation.

Kernel Design in AD:

Most NS signals are generated from stochastic systems and hence show certain trends that govern their underlying time-varying behavior and there existed a need for an automatic discrimination system for such signals, in order to facilitate an optimal characterization for discerning between different classes of data. In this context, two distinct characterization schemes that exploits the kernel trick were proposed.

The first contribution addresses the issue associated with design of optimal TF kernel formulation. Accordingly, AD-based modifications on two existing smoothing kernels (Choi-Williams and Margenau-Hill) were proposed. The evaluation of proposed kernels validated the improved localization and resolution abilities. The introduced kernels provided reduction of the cross-terms in order to obtain a reasonable trade-off between visualization and characterization for efficient feature extraction.

The second contribution involved efficient incorporation of the non-linear mapping during

the TF representation stage using the concept of kernel machines that facilitates extraction of robust features. Further, an investigation on the usability of cross-terms generated when the signal is bilinearly transformed in the TF domain.

8.3 Practical Contributions

AD Analysis of Gait Pattern:

Irrespective of the existence of many schemes, very few schemes that target characterization of NS gait data are available. In this work, an automated altered gait recognition and discrimination scheme has been proposed. Further, a novel feature extraction algorithm for discriminating the ALS pattern from normal gait data (among healthy and ALS subjects) was proposed. A rigorous validation has been performed and the obtained maximum overall accuracy of 93.1% proves the potential of the methodology.

AD Analysis of Pathological Speech:

A detailed investigation revealing the usability of AD for understanding the underlying dynamics of speech disorders has been presented. The proposed AD-based feature extraction approach for quantifying the pathological speech signals offered comparable and in certain cases better performance characteristics for discrimination between the normal and pathological speech. This chapter introduced a high performance speech pathology characterization scheme that provided an overall accuracy of 96.8%.

8.4 Challenges

In most of the available time-frequency based approaches, there exist certain other limitations whenever a scheme for automated signal detection is proposed:

1. The choice of a kernel function is still tricky for a given signal, since the time-varying characteristics might carry a combination of different available kernels. In order to define an adaptive kernel, in the true sense, a hybrid kernel (eg: dictionary-based) approach that exploits the collective advantages of different available deterministic models might be required.

2. For the case of RS-based Fourier transforms, there is yet no theoretical formulation available in regards to the choice of q. Hence, choice of q can affect the amount of spurious signal concentration in the representation.

3. Although it is common to work with a pre-specified number of pattern classes, many problems do exist where the number of classes is not known a priori. This class refers to unsupervised pattern classification.

4. Non-availability of a significant biological marker in order to distinguish between the classifying signal and arriving at a diagnosis regarding the patient: the connection between the two steps may not always be direct. In other words, a pattern classification method may facilitate the labeling of a given signal as being a member of a particular class; arriving at a diagnosis regarding the condition of the patient will most likely require the analysis of certain other clinical information.

For Biomedicine application, the limitation associated with available approaches is twofold:

- non-availability of a metric that provides a reasonable approximation to both the quantitative and the qualitative measures to facilitate diagnosis regarding the patient condition i.e. the connection between the two measures may not always be direct.

- non-availability of a robust scheme that provides reasonably similar performance for a given class of signals, and scope for further work in TF design for signal classification applications when the signal-to-noise-ratio is low.

8.5 Future Direction

- Investigation on generated terms on the TF plane can be extended further to get a better understanding of those cross-term components which appear along with the auto-term frequencies.

- The cross-term research can be used for adaptively building component interactions dictionary, which can further be used to discriminate between classes of multicomponent signals that differ only in certain interactions.

- Although the potential to use machine learning principles for kernel design has been analyzed, additional investigation can facilitate development of more optimized kernel models for better adaptive signal characterization.

- The practical usability of AD-based techniques can be extended for other disorders for efficient diagnosis applications. A direct outcome of such automated analysis is that overall health-care costs incurred due to aging, unnecessary hospital visits and monitoring for long-term care will be reduced greatly.

List of Publications

Published in Refereed Journals and Proceedings

1. L. Sugavaneswaran, K. Umapathy and S. Krishnan, "Ambiguity domain-based identification of altered gait pattern in ALS disorder", J Neural Eng., vol. 9, no. 4, June 25, 2012.

2. L. Sugavaneswaran, S. Xie, K. Umapathy and S. Krishnan, "Time-Frequency Analysis via Ramanujan Sums", Signal Processing Letters, IEEE , vol. 19, no. 6, pp. 352-355, June 2012, doi: 10.1109/LSP.2012.2194142.

3. B. Ghoraani, K. Umapathy, L. Sugavaneswaran and S. Krishnan, "Pathological speech signal analysis using time-frequency approaches", Crit. Rev. Biomed Eng., vol. 40, no. 1, pp. 63-95, 2012.

4. L. Sugavaneswaran, K. Umapathy and S. Krishnan, "Discriminative time-frequency kernels for gait analysis for amyotrophic lateral sclerosis", Engineering in Medicine and Biology Society (EMBC), 2011 Annual International Conference of the IEEE, pp. 2683-2686, Aug. 30, 2011 - Sept. 3, 2011, doi: 10.1109/IEMBS.2011.6090737.

5. M. Z. C. Azemin, D. K. Kumar, L. Sugavaneswaran and S. Krishnan, "Supervised retinal biometrics in different lighting conditions", Engineering in Medicine and Biology Society (EMBC), 2011 Annual International Conference of the IEEE, pp. 3971-3974, Aug. 30, 2011 - Sept. 3, 2011, doi: 10.1109/IEMBS.2011.6090986.

6. L. Sugavaneswaran, K. Umapathy and S. Krishnan, "Exploiting the ambiguity domain for non-stationary biomedical signal classification", Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 1934-1937, Aug. 31, 2010 - Sept. 4, 2010, doi: 10.1109/IEMBS.2010.5627723.

Submitted to or to be submitted to Refereed Journals

1. L. Sugavaneswaran, K. Umapathy and S Krishnan, "Wavelet-Based Characterization of Gait Signal for Neurological Abnormalities", Submitted to Medical Engineering and Physics.

2. L. Sugavaneswaran, K. Umapathy and S Krishnan, "1D and 2D Ramanujan Sums", Submitted to IEEE Trans. on Signal Processing.

3. L. Sugavaneswaran, K. Umapathy and S Krishnan, "Signal Specific Kernel Design in Ambiguity Domain", In Preparation.

Bibliography

[1] T. F. Novacheck. The Biomechanics of running. *Gait and Posture*, 7:77–95, 1998.

[2] H. Zhou, A. Sadka, and R. M. Jiang. Feature extraction for speech and music discrimination. *Content-Based Multimedia Indexing, 2008. CBMI 2008. International Workshop on*, pages 170–173, Jun. 2008.

[3] D. Kim, S-Y. Lee, and S-I. Amari. Representative and discriminant feature extraction based on nmf for emotion recognition in speech. *Proceedings of the 16th International Conference on Neural Information Processing: Part I*, pages 649–656, 2009.

[4] Y. P. Meau, F. Ibrahim, S. A. L. Narainasamy, and R. Omar. Intelligent classification of electrocardiogram (ECG) signal using extended Kalman filter (EKF) based neuro fuzzy system. *Comput. Methods Prog. Biomed.*, 82(2):157–168, May 2006.

[5] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, California, 1990.

[6] K. H. Ruhm. *Deterministic, Non-deterministic Signals;*. Internet Portal "Measurement Science and Technology", 2008.

[7] A. V. Oppenheim and R. W. Schafer. *Digital Signal Processing*. Prentice Hall, 1975.

[8] Massachusetts eye and ear infirmary voice disorders database, version 1.03 (CDROM), 1994.

[9] D. Vakman. On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency. *Signal Processing, IEEE Transactions on*, 44(4):791–797, Apr. 1996.

[10] G. Zhou, J. H. L. Hansen, and J. F. Kaiser. Classification of speech under stress based on features derived from the non-linear Teager energy operator. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1:549–552, May 1998.

[11] R. R. Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5:337–358, 1995.

[12] Y. Zhu and S. Wee. Kalman smoother and its application in analysis of snoring sounds for the diagnosis of obstructive sleep apnea. *World Congress on Medical Physics and Biomedical Engineering 2006 IFMBE Proceedings*, 14(8):1041–1044, 2007.

[13] J. Wei, S. G. Kong, and G. D. Peterson. ECG signal classification using block-based neural networks. *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, 1:326–331, Jul. 2005.

[14] L. Cohen. Time-frequency distributions-a review. *Proceedings of the IEEE*, 77(7):941–981, Jul. 1989.

[15] J. Jeong and W. J. Williams. Kernel design for reduced interference distributions. *IEEE Trans. Sig. Proc.*, 40(2):402–412, Feb. 1992.

[16] B. G. Ferguson and B. G. Quinn. Application of the short-time Fourier transform and the WignerVille distribution to the acoustic localization of aircraft. *Journal of the Acoustical Society of America*, 96(2):821–827, 1994.

[17] H.-I. Choi and W. J. Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(6):862–871, Jun. 1989.

[18] G. Zhenyu, L.-G. Durand, and H.C. Lee. The time-frequency distributions of nonstationary signals based on a bessel kernel. *Signal Processing, IEEE Transactions on*, 42(7):1700–1707, Jul. 1994.

[19] S. B. Hearon and M. G. Amin. Minimum-variance time-frequency distribution kernels. *Signal Processing, IEEE Transactions on*, 43(5):1258–1262, May 1995.

[20] A. H. Costa and G. F. Boudreau-Bartels. Design of time-frequency representations using a multiform, tiltable exponential kernel. *Signal Processing, IEEE Transactions on*, 43(10):2283–2301, Oct. 1995.

[21] M. G. Amin, G. T. Venkatesan, and J. F. Carroll. A constrained weighted least squares approach for time-frequency distribution kernel design. *Signal Processing, IEEE Transactions on*, 44(5):1111–1123, May 1996.

[22] K. Jemili and J. J. Westerkamp. A kernel based system for the estimation of non-stationary signals. *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, 5:3423–3426, May 1995.

[23] M. E. Tagluk, M. J. English, and R. Vincent. Diagnostic application of a universal kernel to time-frequency energy distribution of ECG signals. *Computers in Cardiology*, pages 395–398, Sep. 1997.

[24] L. M. Khadra, J. A. Draidi, M. A. Khasawneh, and M. M. Ibrahim. Time-frequency distributions based on generalized cone-shaped kernels for the representation of nonstationary signals. *Journal of the Franklin Institute*, 335(5):915–928, 1998.

[25] Y. Noguchi, E. Kashiwagi, K. Watanabe, F. Matsumoto, and S. Sugimoto. Time-frequency analysis of a noisy ultrasound Doppler signal with a 2nd figure eight kernel. *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 2:1842–1845, 2001.

[26] A. Belouchrani and M. Cheriet. On the use of a new compact support kernel in time

frequency analysis. *Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*, pages 333–336, 2001.

[27] D. L. Jones and R. G. Baraniuk. An adaptive optimal-kernel time-frequency representation. *Signal Processing, IEEE Transactions on*, 43(10):2361–2371, Oct. 1995.

[28] X. Liao and L. Tao. A class of novel kernel functions. *Proceedings of the 9th International Conference on Intelligent Data Engineering and Automated Learning*, pages 188–192, 2008.

[29] J. George and K. Rajeev. Sinc-Cauchy hybrid wavelet kernel for support vector machines. *IEEE Workshop on Machine Learning for Signal Processing (MLSP 2008)*, pages 356 –361, Oct. 2008.

[30] L. FridtjofWisur-Olsen and R.G. Baraniuk. Optimal phase kernels for time-frequency analysis. *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP 1996)*, 3:1419 –1422, May 1996.

[31] B. Ghoraani and S. Krishnan. Quantification and localization of features in time-frequency plane. *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2008)*, pages 1207–1210, May 2008.

[32] E. Pekalska and B. Haasdonk. Kernel discriminant analysis for positive definite and indefinite kernels. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(6):1017 –1032, Jun. 2009.

[33] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif. Techniques to obtain good resolution and concentrated time-frequency distributions: a review. *EURASIP J. Adv. Signal Process*, 2009:27:1–27:43, Jan. 2009.

[34] E. Sejdic, I. Djurovic, and J. Jiang. Timefrequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19(1):153–183, 2009.

[35] M. F. Kaleem, B. Ghoraani, A. Guergachi, and S. Krishnan. Telephone-quality pathological speech classification using empirical mode decomposition. *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pages 7095–7098, Sep. 2011.

[36] S. Mallat. *A wavelet tour of signal processing.* Academic Press, 1998.

[37] F. Hlawatsch and G.F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *Signal Processing Magazine, IEEE*, 9(2):21–67, Apr. 1992.

[38] B. Boashash. Time-frequency signal analysis and processing: A comprehensive reference. *Elsevier*, pages 1207–1210, May 2003.

[39] M. G. Amin, A. Belouchrani, and Y. Zhang. The spatial ambiguity function and its applications. *Signal Processing Letters, IEEE*, 7(6):138–140, Jun. 2000.

[40] Z. Hong and B. Zheng. An adaptive-kernel design method based on ambiguity domain. *Time-Frequency and Time-Scale Analysis, 1998. Proceedings of the IEEE-SP International Symposium on*, pages 197–200, Oct. 1998.

[41] M. G. Amin and A. Belouchrani. Blind source separation using the spatial ambiguity functions. *Time-Frequency and Time-Scale Analysis, 1998. Proceedings of the IEEE-SP International Symposium on*, pages 413–416, Oct. 1998.

[42] S. S. Gorthi and P. Rastogi. Improved high-order ambiguity-function method for the estimation of phase from interferometric fringes. *Optical Letters*, 34(17):2575–2577, 2009.

[43] G. Garcia, T. Ebrahimi, and J. Vesin. Classification of EEG signals in the ambiguity domain for brain-computer interface applications. *14th International Conference on Digital Signal Processing (DSP)*, 1:301–305, 2002.

[44] R. Hamila, J. Astola, F. Alaya Cheikh, M. Gabbouj, and M. Renfors. Teager energy

and the ambiguity function. *Signal Processing, IEEE Transactions on*, 47(1):260–262, Jan. 1999.

[45] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein. Correlative time-frequency analysis and classification of non-stationary random processes. *Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on*, pages 417–420, Oct. 1994.

[46] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[47] M. Planat, Rosu. H, and S. Perrine. Ramanujan sums for signal processing of low-frequency noise. *Phys. Rev. E*, 66, 2002.

[48] S. Samadi, M. O. Ahmad, and M. N. S. Swamy. Ramanujan sums and discrete Fourier transforms. *Signal Processing Letters, IEEE*, 12(4):293–296, Apr. 2005.

[49] E. Cohen. A class of arithmetic functions. *Proc. Nat. Acad. Sci. U.S.A.*, 41:939–944, 1955.

[50] E. Cohen. Representations of even functions. *Duke Math. J.*, 25(3):401–421, 1958.

[51] S. Ramanujan. On certain trigonometric sums and their applications in the theory of numbers. *Transactions of the Cambridge Philosophical Society*, 22:259–276, 1918.

[52] L. T. Mainardi, L. Pattini, and S. Cerutti. Application of the Ramanujan Fourier transform for the analysis of secondary structure content in amino acid sequences. *Methods Inf Med.*, 46(2):126–129, 2007.

[53] M. Lagha and M. Bensebti. Doppler spectrum estimation by Ramanujan-Fourier transform (RFT). *Digital Signal Processing*, 19(5):843–851, Sep. 2009.

[54] L. T. Mainardi, M. Bertinelli, and R. Sassi. Analysis of T-wave alternans using the Ramanujan transform. *Computers in cardiology*, 35:605–608, 2008.

[55] J. W Cooley, P. Lewis, and P. Welch. The finite Fourier transform. *IEEE Trans. Audio*

Electroacoustics, 17:77–85, 1969.

[56] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83, Jan. 1978.

[57] M. T. Apostol. *Introduction to Analytic Number Theory*. Springer, 1976.

[58] D. S. Mitrinovic and J. Sandor. *Handbook of Number Theory*. Springer, 1995.

[59] R. D. Carmichael. Expansions of arithmetical functions in infinite series. *Proc. London Math Soc.*, 34, 1932.

[60] A. Hildebrand, W. Schwarz, and J. Spilker. Still another proof of Parseval's equation for almost-even arithmetical functions. *Aequationes Mathematicae*, 35:132–139, 1988. 10.1007/BF01830940.

[61] W. A. Sethares and T. W. Staley. Periodicity Transforms. *Signal Processing, IEEE Transactions on*, 47(11):2953–2964, Nov. 1999.

[62] P. Flandrin. Some features of time-frequency representations of multicomponent signals. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, 9:266–269, Mar. 1984.

[63] D. A. Lyon. The discrete Fourier transform, part 4: Spectral leakage. *Journal of Object Technology*, 8(7):23–34, 2009.

[64] C. E. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37(1):10–21, Jan. 1949.

[65] N. Hurley and S. Rickard. Comparing measures of sparsity. *Machine Learning for Signal Processing (MLSP), IEEE Workshop on*, pages 55–60, Oct. 2008.

[66] I. Daubechies and T. Paul. Time-frequency localisation operators-a geometric phase space approach: Ii. the use of dilations. *Inverse Problems*, 4(3):661, 1988.

[67] F. Hlawatschal, T. G. Manickamb, R. L. Urbanke, and W. Jonesd. Smoothed pseudo-Wigner distribution, Choi-Williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison. *Signal Processing*, 43(2):149–168, 1995.

[68] G. Xiong, H-C. Zhao, L-J. Wang, and J-B. Liu. The wide sense time-frequency representation based on the wavelet and its relation with Cohen's class. *Microwave and Millimeter Wave Technology, 2004. ICMMT 4th International Conference on, Proceedings*, pages 677–680, Aug. 2004.

[69] T. Joachims. Making large-scale SVM learning practical. chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999.

[70] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2:265–292, Dec. 2001.

[71] A. Anand and P. N. Suganthan. Multiclass cancer classification by support vector machines with class-wise optimized genes and probability estimates. *Journal of theoretical biology*, 259(3):533–540, Aug. 2009.

[72] P. Honeine, C. Richard, P. Flandrin, and J.-B. Pothin. Optimal selection of time-frequency representations for signal classification: a kernel-target alignment approach. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 3:III, May 2006.

[73] P. Honeine, C. Richard, and P. Flandrin. Time-frequency learning machines. *Signal Processing, IEEE Transactions on*, 55(7):3930–3936, Jul. 2007.

[74] H. Amoud, P. Honeine, C. Richard, P. Borgnat, and P. Flandrin. Time-frequency learning machines for nonstationarity detection using surrogates. *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on*, pages 565–568, Sep. 2009.

[75] M. Davy, A. Gretton, A. Doucet, and P.J.W. Rayner. Optimized support vector machines for nonstationary signal classification. *Signal Processing Letters, IEEE*, 9(12):442–445, Dec. 2002.

[76] A. Rakotomamonjy, M. Xavier, and S. Canu. Non-parametric regression with wavelet kernels: Research articles. *Appl. Stoch. Model. Bus. Ind.*, 21(2):153–163, Mar. 2005.

[77] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Phil. Trans. R. Soc. Lond. A*, 209:415–446, 1909.

[78] M. Davy, A. Gretton, A. Doucet, and P. J. W. Rayner. Optimized support vector machines for nonstationary signal classification. *IEEE Signal Processing Letters*, pages 9–12, 2002.

[79] A. Rakotomamonjy, M. Xavier, and C. Stéphane. Non-parametric regression with wavelet kernels: Research articles. *Appl. Stoch. Model. Bus. Ind.*, 21:153–163, Mar. 2005.

[80] T. S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. *Proceedings of the 1998 conference on Advances in neural information processing systems II*, pages 487–493, 1999.

[81] T. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. *In Advances in Neural Information Processing Systems 11*, pages 487–493, 1998.

[82] R. Sundberg. Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, 1:49–58, 1974.

[83] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[84] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Process-*

ing, 43(5):1068–1089, May 1995.

[85] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Wadsworth and Brooks, Monterey, CA, 1984.

[86] S. Naoki and R. R. Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5:337–358, 1995.

[87] H. Mitsumoto and F. H. Norris. *Amyotrophic Lateral Sclerosis: A Comphrehensive Guide to Management.* Demos Publications, New York, 1994.

[88] S. McClean, P. Millard, E. El-Darzi, and C. D. Nugent. *Intelligent Patient Management.* Studies in Computational Intelligence. Springer, 2009.

[89] L Sugavaneswaran, K Umapathy, and S Krishnan. Ambiguity domain-based identification of altered gait pattern in als disorder. *Journal of Neural Engineering*, 9(4):046004, 2012.

[90] H.G. Kang and J.B. Dingwell. Separating the effects of age and walking speed on gait variability. *Gait and Posture*, 27(4):572–577, 2008. cited By (since 1996) 61.

[91] J. P. Ferreira, M. M. Crisstomo, and A. P. Coimbra. Human gait acquisition and characterization. *IEEE T. Instrumentation and Measurement*, pages 2979–2988, 2009.

[92] G. A. Cavagna, P. A. Willems, and N. C. Heglund. The role of gravity in human walking: pendular energy exchange, external work and optimal speed. *J. Physiol.*, 528(Pt 3), 2000.

[93] T. Oberg, A. Karsznia, and K. Oberg. Basic gait parameters: reference data for normal subjects, 10-79 years of age. *J. Rehabil. Res. Dev.*, 30:210–223, 1993.

[94] S. Dutta, A. Chatterjee, and S. Munshi. An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification. *Expert Systems*, 26(2):202–217, 2009.

[95] H. Liu, Y. Cao, and Z. Wang. A novel algorithm of gait recognition. *Wireless Communications Signal Processing (WCSP), International Conference on*, pages 1–5, 2009.

[96] R. K. Begg, M. Palaniswami, and B. Owen. Support vector machines for automated gait classification. *Biomedical Engineering, IEEE Transactions on*, 52(5):828 –838, May 2005.

[97] Y. Wu and S. Krishnan. Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis. *Medical and Biological Engineering and Computing*, 47(11):1165–1171, 2009.

[98] R. H. Brown. Amyotrophic lateral sclerosis: Insights from genetics. *Arch Neurol.*, 54(10):1246–1250, 1997.

[99] B. J. Goldfarb and S. R. Simon. Gait patterns in patients with amyotrophic lateral sclerosis. *Arch Phys Med Rehabil.*, 65(2):61–65, 1984.

[100] J. M. Hausdorff, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger. Is walking a random walk? evidence for long-range correlations in stride interval of human gait. *J Appl Physiol*, 78:349–358, 1995.

[101] A. R. Meyer, M. Wang, P. A. Smith, and G. F. Harris. Modeling initial contact dynamics during ambulation with dynamic simulation. *Medical and Biological Engineering and Computing*, 45(4):387–394, 2007.

[102] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *J Appl Physiol*, 88:2045–2053, 2000.

[103] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C. K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. *J Appl Physiol*, 82:262–269, 1997.

[104] J. M. Hausdorff, M. E. Cudkowicz, R. Firtion, J. Y. Wei, and A. L. Goldberger. Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. *Movement Disorders*, 13(3):428–437, 1998.

[105] M. Sekine, T. Tamura, M. Akay, T. Fujimoto, T. Togawa, and Y. Fukui. Discrimination of walking patterns using wavelet-based fractal analysis. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(3):188–196, Sep. 2002.

[106] M. Akay, M. Sekine, T. Tamura, Y. Higashi, and T. Fujimoto. Fractal dynamics of body motion in post-stroke hemiplegic patients during walking. *Journal of Neural Engineering*, 1(2):111, 2004.

[107] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. *In Audio-and video-based biometric person authentication: first International Conference, AVBPA'97*, pages 95–102, 1998.

[108] R. Begg and J. Kamruzzaman. A comparison of neural networks and support vector machines for recognizing young-old gait patterns. *Conference on Convergent Technologies for Asia-Pacific Region (TENCON 2003)*, 1:354–358, 2003.

[109] M. Kohle, D. Merkl, and J. Kastner. Clinical gait analysis by neural networks: issues and experiences. *Computer-Based Medical Systems, 1997. Proceedings., Tenth IEEE Symposium on*, pages 138–143, Jun. 1997.

[110] W-L. Wu, F-C. Su, Y-M. Cheng, and Y-L. Chou. Potential of the genetic algorithm neural network in the assessment of gait patterns in ankle arthrodesis. *Annals of Biomedical Engineering*, 29(1):83–91, 2001.

[111] J. G. Barton and A. Lees. An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. *Gait and Posture*, 5(1):28–33, 1997.

[112] M. Köhle and D. Merkl. Analyzing human gait patterns for malfunction detection. *SAC '00: Proceedings of the 2000 ACM symposium on Applied computing*, pages 41–45, 2000.

[113] J. L. Geisheimer, W. S. Marshall, and E. Greneker. A continuous-wave (CW) radar for gait analysis. *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, 1:834–838, 2001.

[114] J. Li and H. Ling. Application of adaptive chirplet representation for ISAR feature extraction from targets with rotating parts. *Radar, Sonar and Navigation, IEE Proceedings -*, 150(4):284–291, Aug. 2003.

[115] P. Setlur, M. Amin, and F. Ahmad. Analysis of micro-Doppler signals using linear FM basis decomposition. *Proceedings - SPIE The International Society For Optical Engineering*, 6210(1):6210–6221, 2006.

[116] V. C. Chen. Spatial and temporal independent component analysis of micro-Doppler features. *Radar Conference, 2005 IEEE International*, pages 348–353, May 2005.

[117] N. V. Boulgouris and Z. X. Chi. Gait recognition using Radon transform and linear discriminant analysis. *Image Processing, IEEE Transactions on*, 16(3):731–740, Mar. 2007.

[118] Y. Wu and S. Krishnan. Statistical analysis of gait rhythm in patients with Parkinson's disease. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(2):150–158, Apr. 2010.

[119] R. C. Wagenaar and R. E. A. Van Emmerik. Dynamics of movement disorders. *Human Movement Science*, 15(2):161–175, 1996.

[120] G. Yogev, M. Plotnik, C. Peretz, N. Giladi, and J. M. Hausdorff. Gait asymmetry in patients with parkinsons disease and elderly fallers: when does the bilateral

coordination of gait require attention? *Experimental Brain Research*, 177:336–346, 2007.

[121] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank. Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology. *Biomedical Engineering, IEEE Transactions on*, 47(6):773 –783, Jun. 2000.

[122] J. M. Hausdorff, Z. Ladin, and J. Y. Wei. Footswitch system for measurement of the temporal parameters of gait. *J. Biomech*, 28(3):347–351, 1995.

[123] *Amyotrophic Lateral Sclerosis (ALS) Fact Sheet*. NINDS.

[124] P. Matja. The dynamics of human gait. *European Journal of Physics*, 26(3):525–534, 2005.

[125] G. P. Zhang. Neural networks for classification: a survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews) In Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 30(4):451–462, 2000.

[126] W. H. Rogers and T. J. Wagner. A finite sample distribution-free performance bound for local discrimination rules. *The Annals of Statistics*, 6(3):pp. 506–514, May 1978.

[127] Y. Wu and L. Shi. Analysis of altered gait cycle duration in amyotrophic lateral sclerosis based on nonparametric probability density function estimation. *Medical Engineering and Physics*, 33(3):347–355, 2011.

[128] M. Banaie, M. Pooyan, and M. Mikaili. Introduction and application of an automatic gait recognition method to diagnose movement disorders that arose of similar causes. *Expert Systems with Applications*, 38(6):7359–7363, 2011.

[129] M. A. Little, P. E. Mcsharry, S. J. Roberts, D. A. Costello, and M. M. Irene. Exploiting

non-linear recurrence and fractal scaling properties for voice disorder detection, 2007.

[130] M. de Oliveira Rosa, J.C. Pereira, and M. Grellet. Adaptive estimation of residue signal for voice pathology diagnosis. *Biomedical Engineering, IEEE Transactions on*, 47(1):96–104, Jan. 2000.

[131] A. A. Dibazar, S. Narayanan, and T. W. Berger. Feature analysis for automatic detection of pathological speech. *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, 1:182–183, 2002.

[132] R. J. Moran, R. B. Reilly, P. de Chazal, and P. D. Lacy. Telephony-based voice pathology assessment using automated speech analysis. *Biomedical Engineering, IEEE Transactions on*, 53(3):468–477, Mar. 2006.

[133] L. Salhi, M. Talbi, and A. Cherif. Voice disorders identification using hybrid approach: Wavelet analysis and multilayer neural networks. *World Academy of Science, Engineering and Technology*, 2008.