Passive Sonar Signal Detection and Classification

By

Mohtashim Baqar

Registration No.: NUST201260533MPNEC45312F

Thesis Supervisor: **Cdr. Dr. Syed Sajjad Haider Zaidi PN**

A THESIS

Submitted to
Department of Electrical and Power Engineering,
Pakistan Navy Engineering College,
National University of Sciences and Technology
in partial fulfilment of the requirements
for the degree of

Electrical Engineering (Communication) - Master of Science

June 2016

Abstract

Sound Navigation and Ranging (sonar) systems have long been employed for acoustic signal acquisition and processing in underwater environment. Since 20^{th} century sonar systems have been in use, though major developments in this domain were made post world war II [56]. Further, major areas pertaining to research in this domain have been the development of efficient detection & classification systems, modelling of underwater environment, tracking and telecommunications. Sonar operates in two modes, namely, active and passive. Both active and passive sonar systems are widely deployed in military applications though they have also been used in scientific and commercial applications.

Signal detection and classification in underwater environment has been the challenge that researchers are fazed with for years primarily due to the non-linear mixing of the noise producing sources and non-stationary (statistical properties of the signal varies randomly if it is observed for longer periods) nature of the underwater environment. Problem at hand in this work is the development of a processing system capable of detecting and classifying objects based on noise radiated by underwater sources. Overall proposed model comprised of two sub-modules; a front-end unit and a back-end unit. The front-end unit is used for extracting distinguishing feature set by employing various detection techniques whereas the second sub-module, the back-end unit, is used for providing automated & efficient signal classification by employing template matching and machine learning techniques. For detection, wavelet analysis (daubechies and symlets), classical signal detection approaches for sonar systems, namely, detection envelope modulation on noise (DEMON) and low frequency analysis & ranging (LOFAR) have been used. Besides, renowned speech signal processing techniques have also been employed for feature extraction, namely, linear predictive analysis (LPA), linear predictive cepstral coefficient (LPCC), mel-frequency cepstral coefficient (MFCC), perceptual linear prediction (PLP/ BFCC) and gammatone cepstral coefficient (GTCC). Further, for the purpose of classification, variants of neural networks (NN) and dynamic time warping (DTW)

have been used. Classifiers include, multilayer feed-forward neural network (MFNN), variable learning rate neural network (VLR-NN), radial-basis function neural network (RBF-NN) and dynamic time warping (DTW). Further, to make a computationally low cost system, dimensions of the feature set were reduced using principal component analysis (PCA) and linear discriminant analysis (LDA). Effects of dimensionality reduction were observed on classification rates. In addition, two relatively newer approaches for feature learning and classification have also been used i.e. convolutional neural network (CNN) and multi-linear principal component analysis (MPCA), respectively. Justification for inclusion of the lateral two approaches has been their effectiveness in problems related to detection and classification of tensor objects. Further in light of recent and latest available studies, convolutional neural networks (CNNs) have worked well with greater efficiency in speech classification problems. Moreover, amongst all the applied techniques, the lateral two have produced best classification accuracies i.e. up to 99.4%. Also, a graphical user interface (GUI) has been developed for performing LOFAR and DEMON analysis on recorded and live streams. All scripts have been written and simulations have been done in MATLAB. In this study, two datasets have been used for evaluating performance of the aforementioned detection and classification schemes i.e. a raw dataset acquired via passive sonar platform, having samples belonging to 4 distinct classes of ships and a synthetic dataset, taken from a database [79], having samples belonging to 20 different classes of underwater objects i.e. sea species and man-made objects. Further, the system was tested under noisy conditions at different levels of signal-to-noise (SNR) ratio i.e. $-20, -10, 0, 10, 20$. Noisy samples were generated via adding standard normal distributed synthetic noise to the source samples i.e. additive white Gaussian noise (AWGN). Results obtained have shown good recognition rates and a lot of promise.

This thesis is dedicated to the memory of my mother and my brother

Late. Hasan Fatima d/o Late. Haji Hasan

Late. Murtaza Baqar s/o Naveed Baqar

(Request for Surah Al-Fatiha)

Acknowledgements

First and Foremost, thanks to almighty ALLAH, The most gracious and merciful, Who gave me strength and courage to finish what I set for few years back. It's that courage that made me complete this thesis successfully through every thick and thin. Whatever I have been able to achieve and blessed with in this lifetime is unquantifiable, couldn't thank enough for what ALLAH has blessed me with. I would like to thank and pay gratitude to Masoomeen (A.S), without Their help and blessings nothing would have been possible.

I am grateful and thankful to my late mother, Ms. Hasan Fatima (Late), her struggles and hard work got me to where I am today. The good in me is because of her upbringing. She was my best critic, I profoundly grateful for her candour and love. She is behind all the success and achievements I have had so far. She sacrificed her life for our (her children) better future, even when she wasn't in the best of physical and mental condition. She was and will always be an inspiration to me. May ALLAH bless her with Jannah (amen) and keeps me on right track so that I can live up to her expectations.

I sincerely thank my thesis supervisor, Cdr. Dr. Syed Sajjad Haider Zaidi PN, who has been a great mentor, an inspiration and a father figure to me. With his unconditional support & patience he played a vital role in restoring the lost self belief in me after the death of my ailing mother. I have learned a lot of life lessons from him during the time I have been working with him. One of the few nicest people that I have ever met during this lifetime. To go with being a nice and humble human being, his technical expertise are unparalleled and of equal praise. At times, he flattered me with his profound knowledge, understanding and critical thinking. I wish him all the success here and hereafter.

I thank my family members, my father, Naveed Baqar and my siblings, Anum Baqar, Murtaza Baqar (Late) and Taqi Baqar for their support and motivation during all the

troubled times. They remained patient, encouraged me and supported me so I was able to do things as I wanted to with confidence. They accepted me as I am and stood shoulder to shoulder with me in every test of life. I wish them all the best of health and life. Here, I would like to make two special mentions. First, of my maternal aunt, Ms. Nishat Fatima and second one of my maternal uncle, Mr. Hamood Husnain, whom have been the biggest support system to my family and I owe them & their families a big thank you for being there in all stiff periods of life.

I thank all my teachers, who taught me at different stages from school to university and contributed heavily in making me a better person. Among them two special mentions that I would like to make is of Dr. Abid Karim, an inspirational figure, a kind hearted human being and for being more than just a teacher to me and of Dr. Muhammad Moinuddin, my undergraduate project supervisor, a teacher, above all a very kind and humble human being. His knowledge and humility inspired me to move forward and seek higher education. Someone, who has always been there to help and guide. I am also thankful to the Guidance and Examination Committee (GEC) members, namely, Cdr. Dr. Attaullah Memon PN, Cdr. Dr. Aleem Mushtaq PN and Cdr. Dr. Hammad Raza, for their necessary guidance and support. I thank and acknowledge the Electrical Power Engineering department and people associated with it for providing all the necessary resources, facilities and environment in making this journey a learned one.

At the end, I would like to make few special mentions without whom this section will go incomplete, starting with my elder brother, Engr. Muhammad Adeel, he has been a mentor, a teacher and a brother in true sense to me. Someone I look up to & listen, one of the very few people who has always been there for me whenever I needed. Secondly, Dr. Jawwad Ahmed, a great teacher, an elder brother, who has been one of my biggest support system. Someone, who has always rejuvenated me with words of praise and motivation. Engr. Azeem Aftab, a friend, a colleague and on top of that a very good human being, who has been of great support to me for the past three years and our bond is getting stronger with time. I wish him all the success here & hereafter. Lastly, I would like to thank all my friends, family and colleagues especially Engr. Laeeq Ahmed, Engr. Syed M. Ghazaal Jafri and Engr. Syed Safdar Hussain for being a motivating force that kept me going, cheering me up when I was down and out. I wish them all the best of luck in their future endeavours.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Introduction

Sound Navigation and Ranging (sonar) signal processing system utilizes propagation of sound in water for object detection, classification and communication. Prime objective of the said system is to analyse underwater environment and use it to advantage in navigation, tracking, surveillance, sea tomography and for detecting fishes etc. It has applications in both commercial and scientific world. It has been used to greater extent in military applications for surveillance and tracking purpose. Sonar operations are put into two categories; passive sonar systems and active sonar systems. Passive sonar only listens to the radiating sources in underwater environment whereas active sonar radiates and listens to it's echo to make decisions. They are mostly used in military settings for surveillance and tracking etc. whereas active sonar systems are used in applications related to sea tomography etc. Sonar setup comprised of sensors i.e. hydrophones, with a backhand processing unit that enables efficient detection and classification of objects. Modes of operation of a sonar system are illustrated in figure 1.1.

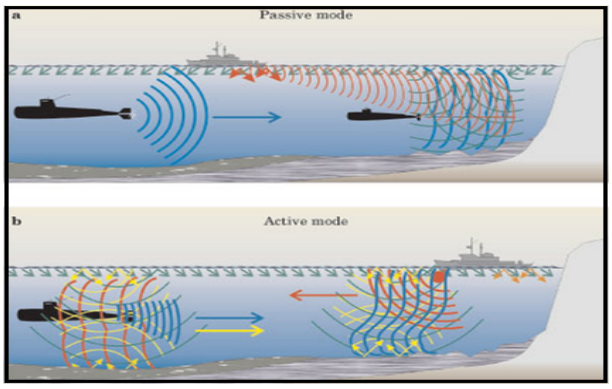

Figure 1.1: Sonar Modes of Operation [56] **(a)** Passive Mode **(b)** Active Mode

As mentioned earlier, mostly passive sonar systems are deployed in military settings. So, that raises the need of a processing system that enables automated decision making for object classification. This thesis aims at building a passive sonar signal detection and classification system for underwater signalling sources. Scope of this work includes performance evaluation of detection and classification schemes for sources in underwater environment.

Passive sonar listens to noise (vibration or sound) radiated by objects. Signals captured using hydrophones are processed for detection and classification of objects. Captured signals can be from sources of different and distinct nature i.e. from a target (Ship, Submarine), self-noise of the sonar platform or the sea ambient noise.

Detection and classification of underwater objects is a complex task. Major factors influencing classification of objects in underwater environment are listed below [40]:

1. Variations in operating conditions of the source or target.

2. Variations in the environmental conditions.

3. Due to the presence of spatially varying clutter of sources.

4. Due to variation in target's shape, orientation and composition.

5. Finding common discriminating feature set for signalling sources acquired using different and spatially dislocated sonar platforms.

6. Bottom features such as coral reefs, sand formations and vegetation plays role in obscuring the target or could confuse the detection process.

Initial step of processing is the estimation of direction of arrival (DOA) of the radiating source. It is usually accomplished using beam-forming techniques [19]. Afterwards, acquired signal from estimated direction is preprocessed to remove artefacts, that is, presence of unwanted signalling components from other sources from neighbouring directions. Proportion of interfering sources in the mixture depends upon the bearings (DOA) resolution. Methods like, independent component analysis etc., are employed to remove unwanted signalling components.

After required signal conditioning and pre-processing, signals are fed to the feature extractor to extract discriminating characteristics of the source. Then, the extracted feature set is passed on to the classification unit for automated recognition. However, signal conditioning and preprocessing is not considered under the scope of the presented work.

To implement detection module, several signal processing techniques have been used, including, wavelet analysis, renowned sonar signal detection techniques i.e. detection of envelope modulation on noise (DEMON) and low frequency analysis & ranging (LOFAR), and some of the most renowned speech signal processing techniques i.e. linear predictive analysis (LPA), linear predictive cepstral coefficient (LPCC), perceptual linear prediction (PLP), mel-frequency cepstral coefficient (MFCC) and gammatone cepstral coefficient (GTCC). For classification module, machine learning techniques have been employed, including, a template classifier i.e. dynamic time warping and variants of neural networks i.e. multilayer feed-forward neural network (MFNN), variable learning rate feed-forward neural network (VLR-NN), radial-basis function neural network (RBF-NN) and convolutional neural network (CNN).

Also, measures have been taken to make an efficient and computationally low-cost detection and classification system. To fulfil the said purpose, dimensions of the feature set have been reduced and effects of it were observed on classification rates. Further, two linear subspace learning approaches i.e. principal component analysis (PCA) and linear discriminant analysis (LDA) and one multi-linear subspace learning approach i.e. multi-linear principal component analysis (MPCA), have been used for dimensionality reduction.

Constraints, including non-stationarity of the underwater environment and signalling mixtures are not considered in this work. The focus is primarily on to build an underwater object detection and classification system to assist the sonar operator.

1.2 Motivation for Proposed Work

Though in-numerous distinct work has been done in area of sonar signal processing for object detection and classification, but still there exist many constraints & challenges

that halts development of a perfect expert system. Few of the challenges that motivated for this proposal are as under:

- Lots of distinct work have been presented in the literature for sonar signal detection and classification but most of it has been targeted to a specific platform or limited to a specific environment. There are no standard go to methods available in literature. Meanwhile, it is very difficult to find correlation between the works available in literature [42].

- Not much thorough work has been done to investigate the performance of detection and classification methods for objects of distinct nature. Majority of the work deals with automated recognition systems for objects of similar nature.

- Most of the literature deals with the use of conventional sonar signal detection techniques, not many thorough studies have been conducted to exploit other acoustic signal detection techniques for object detection in underwater environment.

- Having an efficient detection and classification system been the principal task but computation load of the system is one of the other key factors to be considered because it effects a lot of other system attributes i.e power consumption, memory, delays and processing time etc. Not much work available in literature investigates methods that paves way for development of a computationally low-cost detection and classification system without compromising on recognition rates.

- Multi-linear subspace learning and deep learning approaches have been successfully extended to applications of speech recognition but haven't been tested to classify underwater acoustic transients.

1.3 Research Objectives

Keeping in mind the challenges and constraints towards building an efficient detection and classification system, the objectives of this study are:

- To present a comprehensive study, evaluating performance of various detection and

classification approaches outlined in literature.

- To evaluate performance of the system using samples from distinct underwater acoustic sources obtained via passive sonar platform as well on the synthetic dataset.

- To make the system computationally low-cost, two linear subspace learning methods i.e. principal component analysis (PCA) and linear discriminant analysis (LDA), and one multi-linear subspace learning method i.e. multi-linear principal component analysis (MPCA) will be applied. Moreover, effects of dimensionality reduction on recognition rates will also be observed.

- To evaluate performance of all the applied schemes under the effects of noise i.e. noisy samples will be generated using additive white Gaussian noise (AWGN) at different levels of signal-to-noise ratio (SNR).

- To develop a graphical user interface (GUI), implementing DEMON and LOFAR analysis for recorded as well as real-time acoustic streams.

1.4 Thesis Organization

The thesis comprises of six chapters. **Chapter one** covers the overall introduction to sonar systems, their operations and work undertaken in this study. It includes, motivation for the proposed work as well as the research objectives. Moreover, main objectives of the proposed study are discussed in chapter one. **Chapter two** presents the comprehensive literature review, highlighting major contributions made in this area of research. **Chapter three** describes the proposed detection, classification and dimensionality reduction techniques along with their mathematical models. **Chapter four** gives details of simulation environment and parameters. It also provide detailed summary of simulation results. **Chapter five** provides overall conclusion of the proposed work undertaken in this study. **Chapter six** gives recommendations for future work in this area.

2 Background

Automated recognition of objects in underwater environment is deemed a difficult task because of large variations in both temporal and spectral characteristics of the signals even if they are obtained from one source [56]. Signals acquired via sonar platform comprises of noise radiated from vessels and underwater species i.e. whale, porpoise, mantis shrimp etc. Every signal has its own characteristics that are identified or labelled by human experts either by listening or by investigation of spectrograms of the processed signals. Though, it is not an objective approach for humans to identify tonal characteristics of a source all by themselves. So, this creates need for an automated classification system to reduce the sonar operator's load. For automated recognition, a good feature extraction process is needed and to complement the detection process a strong classifier should be selected. For classification, among many, neural network classifier is one of the most sought after machine learning paradigms that helps in achieving good classification rates even when the problem is highly non-linear. Furthermore, owing to the adaptive nature and parallel processing ability of neural networks, they have been applied in many applications i.e. sonar signal processing [16][30][31], antenna modelling, speech recognition [12][49], facial recognition etc.

Further, studies specifically underlying development and analysis of sonar frameworks are far and few whereas most of them have relied upon synthetic acoustic dataset while some dealt with real-time acoustic samples acquired via sonar platform. Sonar data acquisition is a costly and time taking process which requires a lot of resource, so most of the studies conducted have used synthetic acoustic sample set. Moreover, extraction of optimum feature set and classification are two major components that defines the performance of a sonar signal processing unit. Among the classifiers, neural networks along with Markov chains [9] and its variants, namely, hidden Markov models (HMM) etc., have shown better results. Many research organizations and universities have initiated programs in order to develop sensor based autonomous underwater ve-

hicles (AUV) to serve the needs of military in coastal environments. For example, a program was initiated at Massachusetts Institute of Technology, MIT, to build and test a low Frequency sonar system with the name of Generic Ocean Array Technology Sonar (GOATS) [24]. Passive sonar system comprises of an acoustic receiver that listens to noise radiated by the sources and a processing unit i.e. detector and classifier. Moreover, characteristics of methods used for feature extraction and classification are of particular importance to have good classification accuracies. This chapter presents a comprehensive literature review of the work conducted for signal detection and classification of objects in underwater environment.

Features and Feature Extraction Methods: Characteristics of a source or target lies within it's spectrogram i.e. spectral content. Though, processing systems making use of pure spectral content met little success in the early era of sonar signal processing [80]. In addition, classification of underwater vehicles is much more complicated as compared to surface vehicles. Further, understanding the nature, characteristics of environment and radiating sources is key to achieving good detection and classification rates. Many studies reiterated that characteristics of surroundings and underwater environment are Gaussian in nature while others stated them to be non-Gaussian [83]. Some of the most common ambient noise producing sources with their spectral characteristics as described in [80] [11] are as follows:

1. Seismic Disturbance
2. Biological Organisms Activities
3. Distant Shipping
4. Swirl and Wind
5. Thermal Noise
6. Oceanic Turbulence

Moreover, distant shipping and wind are prime contributors to ambient noise. Level of noise depends upon the distance from the ship and condition of the sea. In deep sea

environment, distant shipping noise is dominant in the frequency regions from about $10-20\ Hz$ to $200-300\ Hz$ whereas wind noise is dominant in the regions from about $200-300\ Hz$ to several tens of KHz. Spectral density of the ambient noise is relatively smooth in the acoustic region.

Radiated Noise: Noise radiated due to machinery and motion of underwater and surface vessels has both narrowband as well as broadband component. Propellers and hydro-dynamic turbulences produces noise covering wider bands. The narrow band components are due to the propulsion system and auxiliary machinery. Moreover, speed of the vessel also contributes to the overall radiated noise and has fair reflection in the spectrum of the source. Major reason of research and development in this area has been due to the strange nature of the underwater environment and its effects on radiated signals. Environment's effect on signal characteristics is described as follows,

1. Rapid change in characteristics of a signal over time and frequency.

2. Variation in the overall energy of the signal due to multipath propagation.

Dominant Noise Generating Mechanisms: Works in [64] [63] [80] have given an excellent description of kind of noises radiated from ships, submarines and vessels in underwater environment. Based on the assessment and analysis performed by authors, these noises were found dominant and categorized as follows:

- **Propeller Cavitation Noise:** The high pressure created on the suction surface of the propeller rotating in water generates cavitation noise. Further, number of blades, blade geometry, propeller's rotation per minute, forward speed of the vehicle and ambient conditions dictates the intensity of cavitation noise. Analysis and calculations have shown that merchant vessels generates the most cavitation noise followed by submarines and warships. Cavitation noise for submarines decreases with submergence.

- **Blade Rate Tonals:** More or less in all marine vessels pusher propellers are used, operating in turbulent and non-uniform wakes. This results in oscillations at multiples of blade-rate frequency.

- **Piston Slap Tonals:** Piston-slap refers to the impact of piston against cylinder

wall. Noise generated due to piston slap depends upon the mounting arrangements and vibration isolation used in ship structures. Though, it is one of the most dominant noise generating source in all diesel and reciprocating compressors used in ships. It is not considered important for slow-speed diesel ships.

- **Gear Noise:** Gear Noise is also one of the common type of noise radiated from marine vessels. Marine engines use reduction gears and transmission error in gears produce gear noise.

- **Injector Noise:** Injector noise is mainly due to the needle settling impact in the fuel injection system. Typical values of noise intensity are in between 50 to 60 dB re 1 $microPA$. It can be a key classification attribute for vessels whose injector noise do not get suppressed due to cavitation.

- **Low Frequency Radiation of Hulls:** Noise due to hull of surface is usually negligible because ocean surface produces negative image of the source within half-wavelength of the hull source, thus almost cancelling it out. Moreover, low frequency hull radiation is an important factor for classification of submarines as the image cancellation is much less and also due to the fact that propeller cavitation noise is also absent.

- **Propeller Speed and No. of Blades:** Propeller's blade causes cavitation noise to modulate at the blade's frequency rate. A standard method to detect such noise is detection envelope modulation on noise (DEMON), the composite noise is filtered using a bandpass filter to isolate the high frequency cavitation noise components. Filtered noise components are passed through a square law demodulator and a low pass filter for detection. Finally spectrum is analysed of the acquired signal for any given target. Number of blades can be determined by dividing the blade frequency with the shaft frequency.

- **Types of Propulsion:** Low speed vehicles have engines with speed up to 400 RPM and produces slap piston tonals in the range of $3 - 9$ Hz. Medium speed vehicles have speeds in the range of 400 RPM to 1000 RPM and the piston slap noise is in the range of 7 to 17 Hz. Moreover, extraction of piston slap tonal can help in classifying objects on the basis of propulsion speed.

- **Injector Noise:** In most marine engines, fuel injector system generates a single frequency component that is at about 1700 Hz and has a broadband spectrum of around 5000 Hz. With spectral analysis total injector noise component can be detected.

In a work, Urick [80] stated that signals radiated from ships can be placed into three categories: machinery signals, propeller signals and hydrodynamic signals. Machinery signals are generated due to vibration of various parts of ship i.e. shafts, armature, gear teeth, turbine blades etc. They produce line spectra; a predominant tonal component with frequency being the fundamental frequency of vibration. In addition, the likes of pumps, pipes and valves contributes to continuous spectra while superimposing the tonal components. Propeller signals are generated outside the hull as a consequence of propeller's movement and by virtue of vehicle movement in water. These cavitation signals form due to the rotation of the propeller. Also, propeller signals also produce a tonal spectra in addition to the continuous spectra of cavitation signals. Lastly, hydrodynamic signals are produced as a result of irregular and fluctuating fluids motion. It consists of Gaussian signals and flow signals generated by the hull of the vessel and the ambient signals in the ocean, respectively. Further, it is quite clear from above discussion that tonal components hold characteristics of the vessels. Some of the tonal components vary with the speed of the engine whereas some stay the same.

Work in [61], presented a study for passive sonar signal detection and classification where a feature set representing four classes of ships were taken into account. An expert system RECTSENSOR was developed to classify objects. Methodology adopted by the system to calculate accurate decision was based on slightly modified version of Dempster-Shafer theory. This study also highlighted major noise generating mechanisms in ships & submarines along with methods to identify key discriminating feature set.

- **Expert System:** The proposed system had three major elements; an inference engine, a knowledge base and a database. The knowledge base has all the rules and regulations relating to the facts about the object and environment. Database used, has been developed according to the context. The inference engine helps in overlooking the reasoning process in conjunction with the database. In the proposed model, the knowledge base was created using PROLOGUE, an object oriented programming language with a unique ability to infer facts from other facts. Using heuristics, PROLOGUE executes the matching process. It attempts to find

out facts that satisfies a particular condition. If it fails, the system back-tracks to the previous fact and tries to prove it again with a new binding. The actual implementation of the expert system took only four vessels into consideration while using nine discriminating features in the feature set. Flow of the system was such that on acquiring signals, it extracts features and store them in a database through a dialogue session. This session could be that the system may ask the user to provide information about the type of propulsion etc. Moreover, the selected type is stored in the database. It is imperative that the performance of the expert system depends upon the feature extraction method as well as the measure of the portion of the total belief which is committed to different classes of vessels by these attributes. An accuracy index is used so that it gives the measure that the direction is towards the correct solution. The accuracy index can hold values between 0.5 and 1. Its value is decided by taking into account some conditional probabilities pertaining to features and targets.

Also, work has also been put in to model cavitation in acoustic environment. In a work [44], authors tried to improve the cavitation model by including the effects of acoustic losses which comes from taking the compressibility of fluid into consideration. Moreover, losses due to heat were also considered. For sonar signal analysis, building a cavitation model is a necessity while working on detection and tracking of objects based on radiated cavitation noise. In another study [17], a model was build to find the effects of masking by shipping and surrounding noises on the sounds generated by the protected animals for communication purpose i.e. whales. As spectra of both comprises low frequency components and size of species makes it nearly impossible to conduct any experiment in captive, controlled and close environment. So, this study gave a qualitative analysis of the noisy effects that shipping and surrounding sounds may have on the sounds produced by the sea specie i.e whale.

In [15], Chin-Hsing Chen applied neural network classifiers to investigate the feasibility of neural networks to problems where the tonal component is varying with respect to the speed of the object.

Apart from the usage of spectral contents as features for classification of objects, studies have also shown the usage of autoregressive models for automated recognition of objects [25]. In a study, autoregressive models were used to classify three types of propulsion systems i.e. high speed diesel, low speed diesel and turbine [44]. Also, there

have been studies to categorize vessels. Mainly, classification is done based on vessel speed, blade-rate of propeller, location of the tonal components of the machinery, injector noise, gear noise and due to the low frequency radiations from the vessel. All of these mentioned features can be extracted from the spectrum of the noise generating object. Many estimation techniques have been proposed exploiting patterns in the spectrum of the objects for efficient detection and classification. Algorithm proposed in [43] used about six parameters, acquired from the spectrum of the source and classification was performed by comparing the extracted feature vector to the ones in the database using euclidean distance as the metric for comparison. Due to the presence of noise, unknown features appear in the spectrum that makes classification difficult or impossible at times. Another method discussed in [74] utilizes power spectrum for distinguishing between four different vessels. Moreover, the method utilizes two-pass split window (TPSW) for estimation of background noise of the sonar platform.

Further, auto-Regressive (AR) models and cepstral coefficients are highly dependent on signal-to-noise ratio (SNR) while spectral components being more robust to variations in signal over time. Moreover, AR models are usually used in stationary conditions or in short time signal processing. features are used as inputs to the classifier. A lot of studies suggested HMM over other classifiers because of its optimum performance and robustness. In an environment prone to noise, radiated signals from a source may suffer degradation and interference, but the spectral features seem to remain unchanged, this highlights the importance of dealing with spectral features for classification. However, in severe noisy conditions, it is difficult to detect main features. Moreover, spectrum may very well be showing wrong amplitudes or positions of features.

For underwater object classification, feature set required to discriminate objects usually comprised of spectral contents of the sources. Spectral densities are more robustness and give better estimates as compared to statistical or parametric modelling of sources. But using only the spectral information obtained from sources with out any transformations can yield false detection and classification. It is because of high variations in spectral characteristics of a signal, mainly due to the presence of ambient and system noise. More or less, methods available for detection and classification takes power spectral density estimates into account but due to the environmental conditions, performance level of the system may degrade substantially. In such cases either some transformation helps in deducing discriminating features or some higher order spectrum can help in target recognition. In a work [84], authors suggested to use higher-order es-

timates in conjunction with the power spectral density estimates for target recognition. This work analysed the advantages and disadvantages of using power spectral estimates and higher-order estimates as features. Moreover, a multilayer neural network was used for classification. Results observed showed that usage of BP neural network and combining the two said information improves the classification rates. In another work [43], an expert system, named "EXPLORE" was developed to identify the types of noise radiated by underwater objects. Set of targets, comprised of surface ships, cargo ships, speed boats, submarines and fishing vessels were taken into account. System used fuzzy logic and clustering for decision making. Moreover, feature vectors were mapped into the Hilbert space which then were used for detection. System was tested with several hundred samples and the results obtained showed that the expert system, EXPLORE, is intelligent in recognizing underwater targets even at very low signal-to-noise ratio (SNR).

Markov models have been very effective in classification problems having to deal with variable length feature set i.e. speech recognition and classification.

Markov Models: Markov models and their variants have shown greater stability in classification problems related to objects in underwater environment. They have proven to be at the optimum in detection/ classification problems related to acoustic signal recognition [60]. Presently, many speech recognition algorithms use HMM as classifier where speech sequences are modelled as states of the HMM classifier with probabilities assigned to each state of the classifier. Moreover, probability densities are associated with acoustic observations. These observations are spectra or cepstra corresponding to several states and state transitions. The states are identified while observing the patterns in the acquired spectra. In speech, a left to right topology is used while a mesh topology can be very useful in sonar signal detection and classification, where transition from one state to any state is possible. The left to right model is known as bakis model whereas a fully connected topology is referred as an ergodic model. Although HMM approach has two major problems and they are,

1. Identification of optimum methodology for feature extraction.

2. Training of model for extracted feature set.

Neural Networks and Their Variants: Neural Networks are widely used in classification problems because of their ability to learn and cluster [41] [28]. Amongst several neural network architectures, radial-basis function neural network (RBF-NN) is extensively used due to its fast convergence and low computation cost. RBF-NN works on the criteria of Euclidean distance measure and is extremely sensitive to the magnitude of feature vector. Two vectors having to be apart in space can produce high Euclidean distance measure whereas in contrast, Hausdorff similarity measure (HSM) is able to combat this separation in space with similar Euclidean measure and discriminate them well. Moreover, Hausdorff similarity measure (HSM) is usually used for two dimensional feature set [8].

Another approach that has been used a lot for underwater object classification is probabilistic neural network (PNN); having one input, one hidden and one output layer. Number of neurons in the input layer depends upon the dimension of the feature vector, number of neurons in the output layer are the same as the number of class labels and neurons in the hidden layer depends upon the factor 'n' whereas the parameter 'n' is chosen, such that, good convergence can be achieved i.e. error function should be at optimum minimal. In PNN, parzen method for PDF estimation is used. At the end of the training mode, each class appears as the center of the Gaussian function i.e. mean of the Gaussian function. The conventional PNN needs an input parameter called the spread value of the parzen window. The variance of the Parzen window is directly proportional to this spread value. The difficulty or disadvantage is the selection of an appropriate spread value. Too small value produces a spiky PDF whereas too large value produces a smooth PDF. To overcome this problem, a new approach of multi-spread PNN was proposed [26], where each class can be assigned a different spread value. This method is useful where the intra-class variance is different in each class. MSP-NN is a neural network technique which estimates the PDF of the training data set using Parzen window while using different spread values for each class. Results obtained with MS-PNN are much better compared to those obtained with PNN.

Another article [81] discussed an approach for feature extraction and classification i.e. short-time Fourier transform (STFT) for feature extraction and finite impulse response neural network (FIR-NN) for classification. A database (courtesy: Defence Research Establishment, Canada) was used for performance evaluation of the proposed system. Apart from neural network classifier, other machine learning methods were also employed with former giving better recognition rates. In a study, Duda [23] used Probabilistic

Neural Network (PNN) for classification of objects based on features extracted using autoregressive model. Performance evaluation of the proposed system was made using a synthetic and a recorded acoustic dataset. Authors in [84] discussed the advantages and disadvantages of keeping both the low frequency as well as the high frequency contents of the acoustic samples as part of the feature vector. Features were classified using a modified artificial neural network via back-propagation algorithm. Two feature sets were fed to the neural network classifier i.e. feature set acquired via autoregressive model whereas second feature set comprised of pure spectral contents of the radiating sources. The modified neural network classifier was referred as the multi-spread probabilistic neural network (MS-PNN). As discussed, noise radiated by marine objects in ocean contains information about their attributes while spectral estimates of these noisy signals are used for the detection and classification of objects. In a work, hydrophones were placed in an ocean at a far-off place from the ships while listening to the radiated noise. Classification of ships was made using a neural network classifier that took spectral estimates as inputs. The study also showed that spectral averaging improves the overall classification rates compared to processing of data in frequency domain for each window. It was also observed that cases when object wasn't changing its speed and operating conditions, the radiated noise has virtually an unchanged statistical pattern and was termed as wide-sense stationary noise [71]. Though, when this sound was received at a distance by a hydrophone, it cannot be considered as stationary any more, due to varying characteristics of the ocean. The variations are very much with respect to time and could also be due to the relative geometry between the ships and sensors [80]. It depends upon factors such as ship speed, distance between ship and sensor, propagation characteristics of ocean and the overall environmental conditions. Signals can also be thought of as stationary when variations are slow and steady while propagation towards the receiver as well as when frames of smaller durations are considered for analysis. For a random noisy signal, but stationary, averaging increase the overall signal-to-noise ratio (SNR) while suppressing the incoherent noise spectra. In a work [74] after data averaging, estimates were used to train a neural classifier. Classifier used back-propagation algorithm for cost minimization while learning rate was also varied adaptively as a function of output error. Hyperbolic tangent function was used as an activation function and as mentioned earlier, signal averaging showed improvements in overall classification results.

In another work [6], a modified version of a multilayer feed-forward neural classifier was used for performing a pattern recognition task i.e. IRIS recognition. The well-known back-propagation algorithm was used [33]. Back-propagation technique is one of

the most vastly used learning mechanisms in supervised neural nets. Based upon error calculations, weights are updated at each layer for each respective node. The process is repeated until the cost function approaches an optimum value. Mean square error (MSE) is usually used as a measure of the global error. In this work, a modified version of MFNN was used, which took into account the energy of error while setting the value of learning rate. Classification results showed improvement compared to results obtained using conventional multilayer feed-forward neural network.

In a work [28], a comprehensive study was made to build a good classification system for temporal signalling sources. Authors have explained as how to extract good signal descriptors as feature vectors for a high performance classification system. Moreover, wavelet-based features were considered to be more superior and powerful in terms of classification accuracy compared to autoregressive coefficients and power spectral estimates. A variety of neural classifiers were tested, evaluated and compared with traditional statistical classification techniques. The focus was on those networks that were less susceptible to noise and were able to time-out irrelevant noisy features. This work took into account two neural network based classifiers. Further, methods to combine several different classifiers were also proposed and evaluated for better classification rates. Performance of the system was evaluated using signals from a dataset, namely, DARPA-I.

In a work [20], authors have developed a preprocessing method to improve the classification results. For classification, artificial neural network was used to classify four classes of ships. Using preprocessed data, classification was improved to an accuracy of 97%.

In [75], a statistical data analysis technique, principal component analysis was used with the frequency domain estimates of the signal and the results were used as inputs to an artificial neural network for classification. Classification results obtained with three different feature transformation techniques were compared, the techniques were; linear principal component analysis (PCA), non-linear principal component analysis (NLPCA) and a neural discriminant analysis technique (NDA). Results showed that feature set obtained using NDA outperformed the ones obtained using PCA and NLPCA in terms of percentage classification. Classifier using features obtained with NDA gave a classification accuracy of 93% using only 3 components while with PCA and NLPCA same accuracy was achieved using as much as 33 components.

Multi-target Detection and Tracking: Many researchers have tried their hands in the field of target tracking, some recent developments include Monte Carlo techniques and the tree search techniques combining the data association and target tracking problem as a unified problem. Algorithms have been developed for single target as well as for multiple target tracking [38]. Some of the work has been extended with the introduction of stacked-based tree search algorithm for multi-target tracking in a highly cluttered environment [65] [66] [50]. Target tracking is an integral part of surveillance and with these techniques available, tracking systems can be made to work efficiently in cluttered and noisy environments. As mentioned, a lot of work has been done for multi-target detection and tracking[27]. Multi-target tracking referred as two problems to be solved together, one is estimation and the other one is data association. It is concerned with the estimation of states for unknown number of targets. Available measurements may be acquired from the clutter or targets of interest. Due to the element of ambiguity, it isn't just an estimation problem. Methods available utilizes measurements which lie nearest to the available set of measurements to make predictions. On the other hand, an extended approach uses probabilistic data association filters, this extended work simply relies upon the probability that a measurement is originated from a particular target. It is worth emphasizing that fundamental difficulty in multi-target detection lies in data association. This area over recent years has gained the attention of a lot of researchers because performance of an algorithm directly effects the performance of the sonar platform. With the technological developments, this data association problem has been formulated as a computing problem [13] [48] [54] [57] and labelled as a NP-Hard problem i.e. the complexity increases exponentially as the number of measurements or scans increases. Earlier, this area wasn't exploited intensively but in lateral stages of 70's a real time algorithm was proposed, namely, Multiple Hypothesis Tracker [62] in which the measurements received were assigned to the initial targets, new targets or false alarms. Hypotheses is used to retain the most likely of targets. Moreover, worry was the elimination of the correct measurement sequences as one of the prime reasons of that may be the weak strength or constant fluctuation of signal of the radiating source of interest. Another way of dealing with the data association problem is to consider it from a probabilistic point of view to calculate the likelihood. But the uncertainty still lies and with the increase in number of parameters to be estimated, the uncertainty grows. If the vector to each target is known, then it is mare an estimation problem and the likelihood function can be easily calculated. One of the methods for calculation is expectation and maximization algorithm. Avitzour [5] saw the first use of this algorithm for multi-target tracking. Authors in [78] [77] used EM algorithm in conjunction with

Kalman filtering for multi-target tracking, the algorithm was referred to as Probabilistic Multi-Hypothesis Tracking (PMHT). The probabilistic approach did not require any assignment of measurements to targets. The measurement assignments were taken as random variables and were estimated jointly with possible target states. Hence, the problem was solved only with the probabilistic approach and wasn't classed as a NP-hard problem.

In another article [72], authors have suggested an optimal technique for target tracking which minimizes the effects of correlation uncertainties in underwater environment and improves the performance of the system substantially compared to some of the standard methods i.e. standard Kalman filter. In case of no correlation uncertainties, filter reduces to a Kalman filter. This method is viable in environment where correlation uncertainties cannot be ignored. The filter formulation has two major key elements; two key tracking loop function and return-to-track correlation. It allows evaluation of system performance under its influence analytically. The suggested filter design showed good results on simulation level but it is yet to be applied on real-time data in noisy environment.

In a work [67], an extension to tree-search based tracking mechanism for multi-static target was presented and the performance of the work was tested on a sonar dataset, namely, SEABAR'07. The tree based search mechanisms were originally introduced in [**?**] and were build on stack algorithm for convolutional decoding. To estimate a target, the tracker navigates a search tree in which each path represents a sequence of states the target goes through. Estimation of a track is done via traversing through only a subset of the tree, the stacked based tracker computes only likely regions of the posteriori probability distribution at each update. Thereby, giving a Bayesian inference solution to the problem. In this piece of work, authors have extended the mono-static stack based tracker to multi-static tracking. The structure of the tree helps in facilitating multiple sources detection with minimal increase in complexity. Results observed showed that the tracker has been able to effectively follow the targets trajectory while the sources were exhibiting non-linear manoeuvres in a cluttered environment.

Source Separation: Usually signals acquired through hydrophones are mixtures coming from separate sources and their separation is one of the integral factors on which the performance of the detection and classification system greatly resides. Due to this,

a lot of work has been done in area of blind source separation for extracting original source from the acquired mixture. In a work [10], a broadband approach to blind source separation for convolutive mixtures based on second-order statistics was proposed. This method avoids the initial problems in the conventional narrowband approaches as it takes both non-whiteness and non-stationarity of the source signal into consideration. Also, a novel method for optimization of the cost function was introduced. This approach allowed rigorous derivations of both novel and conventional algorithms to provide better solution of the internal permutation table. Experimental results showed that this theoretical approach leads to better performance results in reverberant acoustic environment in both frequency and time domain compared to practical Blind Source Separation (BSS) algorithms.

Feature Selection: Feature selection is an essential element when it comes to pattern classification, selecting right features has always been an area of concern. In a work [55], a study was made in order to select features according to maximal statistical dependency based on the criteria of mutual information. Because of the complexity in implementing the maximal dependency condition directly, an equivalent form was derived called minimum redundancy and maximum relevance criterion (mRMR). Then a two stage feature selection algorithm was presented combining mRMR with other feature selectors. This also reduced the computation cost and memory requirements as the most relevant features were selected while making the feature vector size compact. An experimental study was presented to evaluate the performance of the algorithm. For the purpose of classification, three different classifiers were used, namely, nave bayes, support vector machine (SVM) and linear discriminant analysis (LDA). Experiment was conducted using four different data sets, including, both continuous and discrete data sets. Results showed significant improvement in overall classification accuracy.

This chapter presented a literature review related to the area under study. Next chapter discusses each front-end and back-end approach used for signal analysis of the sources under study.

3 Proposed Methodology - Feature Extraction and Classification Techniques

3.1 Introduction

Building blocks of any object recognition system has two main sub-components, a feature extractor and a classifier. In this study, work has been undertaken to develop and evaluate performance of the said building blocks to have an efficient detection and classification system. Figure 3.1 illustrates the overall system model of an underwater acoustic signal detection and classification system.

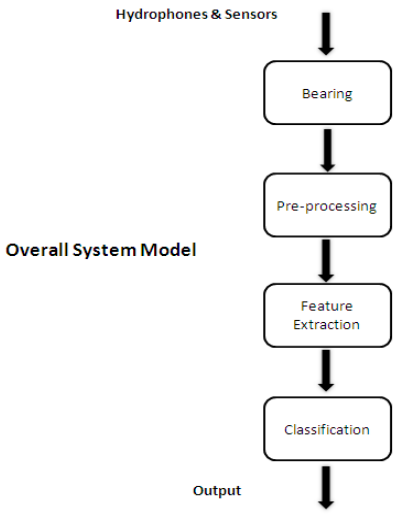

Figure 3.1: Overall Model of Sonar Signal Detection and Classification System

Main objective of feature extraction process is to obtain best discriminating feature set for the acoustic samples under study i.e. representing the characteristics of the samples while reducing the amount of redundancy. Moreover, the acoustic samples were divided into frames of duration up to 25 ms. After framing, each frame is windowed using a suitable windowing function to avoid spectral leakage and generation of intermodulation products. Further, features are extracted for each frame and regarded as feature vector for the frame or at times, vectors of multiple frames are combined to form a single feature vector. Almost in all detection techniques, consecutive frames are overlapped by a percentage, usually 50%. In this study, multiple methods were used for extracting features from the acoustic samples under test. Extracted feature set was fed to multiple back-end modules, that is, classifiers, for automated recognition. Further, the task of a recogniser is to appropriately identify class label for each of the input vector. Moreover, to have good classification results, feature set should comprised of the most discriminating features. Appropriate selection of classifier and good training also leads to good classification results. There are two types of classification techniques used in this study, one is template matching i.e. dynamic time warping (DTW) and other is neural network classifier. In dynamic time warping, feature vector of the acquired sample is matched with the templates of acoustic samples under consideration. The closest found template based on minimum Euclidean distance is the classification result. On other hand, neural networks are modelled such that the input samples are map to their respective class labels.

Moreover, good recognition rate is subject to goodness of the feature set. No matter how strong the classifier is, good classification results cannot be achieved if the input to the classifier does not hold discriminating characteristics of the objects. Various feature extraction and classification techniques have been put to test in this study. All simulation results are discussed in chapter 4. Acoustic profiles were taken from a database [79] as well as from sonar platform to evaluate performance of all detection and classification schemes.

There are two important processing measures that improves the overall detection process. They are briefly discussed below,

Windowing

After framing, first step is to apply a windowing function on each frame as a preprocessing measure. Windowing mitigates spectral leakage and reduce the generation of intermodulation products [58]. Moreover, spectral content of a signal holds characteristics of the source and taking FFT gives us detailed description of the frequency content in the signal. Further, FFT can only take a finite input sequence. Moreover, actual Fourier transform (FFT) assumes signal to be comprised of finite data points and a continuous spectrum, that is, it considers single period of a periodic signal. A periodic signal has periodic footprints in both time and frequency domain, this states that starting of next and ending of previous cycle occurs from the same point in both time and frequency domain. However, at times the ensemble of a signal doesn't comprise one complete cycle or integer multiple of the period of the signal whereas the need of finite set of data for processing needs signal to be truncated. This could cause sharp discontinuities, which may result in a spectrum with different characteristics as opposed to the ones in the original signal. The discontinuities adds up to the spectrum as high frequency components and if those high frequency components are greater than the Nyquist frequency range than they will be aliased in the range of interest and the resulting spectrum is a smeared version of the original signal. It shows as if the energy from one frequency component leaks into another frequency component and termed as spectral leakage. Further, windowing tends to suppress the energy of the samples near discontinuities and thus avoids creation of those unwanted frequency components in the spectrum. Criteria for selecting windowing function is briefly discussed in the following,

1. If an interfering frequency component exists having large magnitude at a distant position from the frequency of interest, then windowing function selected should have a high roll-off rate for the side lobes.

2. If an interfering frequency component having large magnitude exists near the frequency of interest, then windowing function selected should have a low maximum side lobe level.

3. When two or more frequency components lie very near each other and while the time spectral resolution is of key importance then in such cases it is best to choose a smoothing windowing function with a narrow main lobe.

4. If relative or absolute amplitude of the signal is more important than the frequency value then choose a windowing function with a wider main lobe.

5. If the signal has a flat frequency spectrum, choose a uniform window or no window.

6. Hann window gives satisfactory results in almost 95% of the cases. Its main features include, good frequency resolution and low spectral leakage. Moreover, if nature of the signal is not known, Hann windowing function is best to use.

Even when no window is used, signal is multiplied (convolved) with a *sinc* function (rectangular-shaped window) having uniform height in frequency (time) domain. It is done to extract an ensemble from the input stream, so that a discrete sequence can be obtained for processing. No window is often called the uniform or rectangular window because there is an effect of spectral leakage visible in the signals spectrum.

Hamming and Hann are most widely used windowing functions because of their features. Both are sinusoidal in shape and produce wide peak main lobes and low peak side lobes. Moreover, Hann windowing function touches zero at both endpoints eliminating any possible discontinuity in the signal whereas the Hamming window doesn't quite reach the zero level at both discontinuities, thus does a poorer job at completely eliminating discontinuities present in the signal although it does an excellent job in cancelling any side lobe near the main lobe. These windowing functions help in retaining information content as in the original signal while eliminating noise and giving good frequency resolution at same time.

Two-Pass Split Window (TPSW)

Two-Pass Split Window [51] is applied for spectrum smoothing against background noise. The frequency content of the radiated signal consists of two major components. First is a broadband component which has a continuous spectrum i.e. noise and the other is a tonal component, which has a discrete spectra. Mainly, tonal components in the spectrum are the characteristic features of the source. So, extracting these tonal components is key in having good detection and classification results. Further, TPSW works well in extracting the tonal components from the continuous spectrum. Steps of the algorithm are listed below.

1. For a signal $f(x)$, a window is selected, centred on the current sample, k, in time as, $R_k = \{k - M, k - M + 1, \ldots\ldots, k, \ldots\ldots, k + M - 1, k + M\}$. Where, length of the selected window is $2M + 1$.

2. In pass-1, mean for the window centred at k is calculated. It is done for all values of k,
$$\hat{f(k)} = \frac{1}{2M+1} \sum_{i=K-M}^{i=K+M} f(i)$$

3. Further, a clipped sequence, $g(k)$, is formed in order to avoid biasing of the estimates of local mean due to the presence of tonal components,
$$g(k) = \begin{cases} \hat{f(k)} & if, \quad f(k) \leq \alpha \hat{f(k)} \\ f(k) & if, \quad f(k) > \alpha \hat{f(k)} \end{cases} \qquad \text{where } \alpha \text{ is a constant}$$
where, α is a constant. Its typical value is 0.9.

4. Next is pass-2, continuous spectra is again attained by evaluating the local mean of the sequence obtained in pass-1, that is, $g(k)$,
$$\hat{m(k)} = \frac{1}{2M+1} \sum_{i=K-M}^{i=K+M} g(i)$$

On estimation of the broadband component, narrow band components or tonal components can be evaluated by removing the estimated spectrum from the spectrum of original signal, i.e.,
$$h(k) = f(k) - \hat{m(k)} \tag{3.1.0.1}$$
The tonal components thus extracted are normalized to avoid any amplitude discrepancies and also because we are only interested in the patterns present in the spectrum,
$$X = \frac{h}{\| h \|} \tag{3.1.0.2}$$

3.2 Detection Techniques

This section gives detailed insight into all the feature extraction methods used in the implementation of proposed study. Including, wavelet analysis, classical sonar signal detection techniques along with some of the most popular speech signal processing techniques. Details of all the detection schemes are in the following.

3.2.1 Detection Envelope Modulation on Noise (DEMON)

A narrow band signal analysis technique which is heavily used in detection of signals acquired via sonar platform [51]. Moreover, noise radiated by sources have characteristics in the spectral content. Further, DEMON furnish propeller's characteristics i.e. Shaft Rate, Shaft Rotation Frequency, Blade Rate and Number of Blades etc. Results of DEMON analysis are often credible and helps in efficient recognition of underwater object. Figure 3.2 depicts block diagram for implementing classical DEMON analysis.

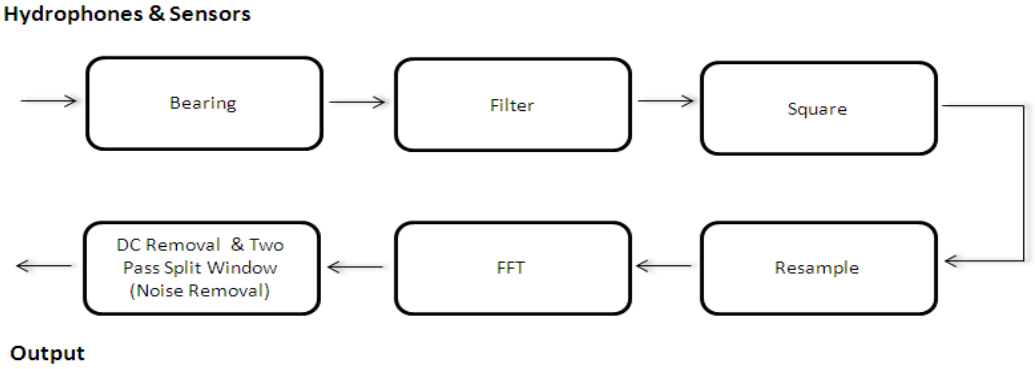

Figure 3.2: Detection Envelope Modulation on Noise (DEMON)

Given the direction of arrival (DOA), acquired signal is fed to a bandpass filter, that is, to select the range of interest while limiting variations in the signal. Moreover, oscillations in acoustic signals range from a few Hz to hundreds of Hz, proper selection of cavitation range is one of the key elements in having good detection. Further, the filtered signal is squared as done in classical envelope detection technique. As this is an era of digital signal processing, so, most of the end-systems are digital and samples at very high rates. But that much sampling rate is useless in underwater environment as most

of the underwater objects have characteristics in low frequency contents. Therefore, after filtering, signal is down sampled according to the observational needs and this increases the coarse resolution of the range of interest. Finally the signal is frequency transformed using Fourier transformation (FFT), so it can be analysed in frequency domain. In addition, TPSW method is applied to the resulting spectra to reduce the background noise. DEMON is usually implemented in two different ways. They are listed and briefly discussed in the following.

Demon - Hilbert Transform

From a real data sequence, Hilbert transform yields a complex signal, also referred to as an analytic signal. The complex signal $x = x_r + i * x_i$ has a real part, x_r, which is the actual data, and an imaginary part, x_i, which comprises information acquired after Hilbert transformation. Imaginary part has version of the original signal, which is 90^o phase shifted i.e. *Cosine*s are transformed to *Sine*s and vice versa. Moreover, the transformed series has same frequency and amplitude content as in the original signal. It also includes phase information which depends on phase of the original sequence. Figure 3.3 illustrates implementation of DEMON using Hilbert transform.

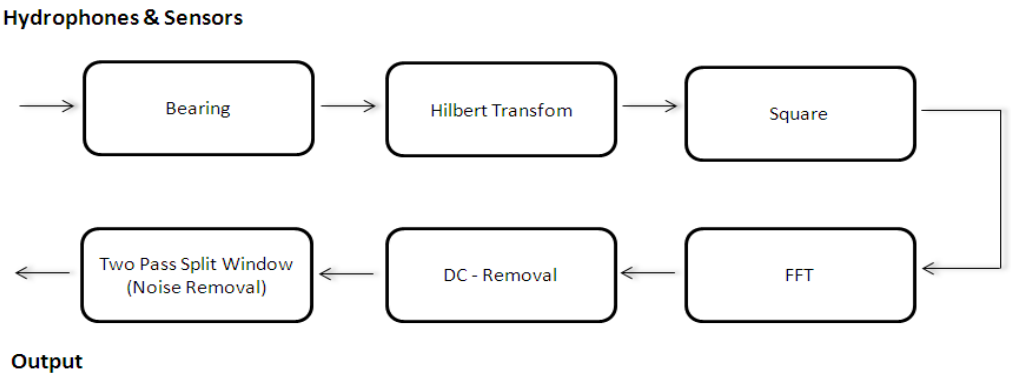

Figure 3.3: DEMON - Hilbert Transform

Hilbert transform is useful in calculating instantaneous attributes of a signal, specially frequency and amplitude. Instantaneous amplitude refers to the amplitude of the Hilbert transformed signal whereas instantaneous frequency is the rate of change of the instantaneous phase of the transformed signal. For a sinusoid, instantaneous frequency

and amplitude remains constant. Instantaneous phase, however, is like a sawtooth, depicting linear phase shift over a single frequency cycle. For a signal having combinations of sinusoids, signal attributes are short term or local i.e. averages not lasting for more than two or three points. For DEMON analysis, each frame is transformed into Hilbert space, transformed signal is squared as done in envelope detection. Further, squared signal is frequency transformed using Fourier transform (FFT) and DC component is removed from the resulting spectra. In addition, spectrum is smoothed using TPSW. Mathematically, Hilbert transform, $\hat{x(t)}$, of a signal, $x(t)$, can be expressed as,

$$\hat{x(t)} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(s)}{t-s} ds \tag{3.2.1.1}$$

where, integral represents the Cauchy principal-value integral.

The reconstruction of the original signal can be achieved using the following formula,

$$x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{x(s)}}{t-s} ds \tag{3.2.1.2}$$

The frequency-domain description of Hilbert transform can be mathematically expressed with the following equations,

$$H(\nu) = -j.sgn(\nu) \tag{3.2.1.3}$$

where,

$$sgn(t) = \begin{cases} -1 & if, \quad t < 0 \\ 1 & if, \quad t > 0 \end{cases}$$

So,

$$\hat{X(\nu)} = -j(sgn(\nu))X(\nu) \tag{3.2.1.4}$$

DEMON - Low Pass Filtering

Another approach to implement DEMON is using a low pass filter. Figure 3.4 shows the overall system model of DEMON implementation via low pass filtering. Moreover, this approach is analogous to amplitude demodulation as the input signal is squared as done in envelope demodulation. Then, the squared signal is fed to a low pass filter to eliminate any unwanted high frequency components. Cut-off frequency of the filter is chosen while keeping in mind the frequency range of interest. Next, square root of the

filtered signal is taken to mitigate the effects of earlier squaring. Further, the resulting signal is frequency transformed using FFT and DC bias is removed from the transformed signal. Lastly, the resulting spectra is passed through TPSW to suppress the effects of background and ambient noise.

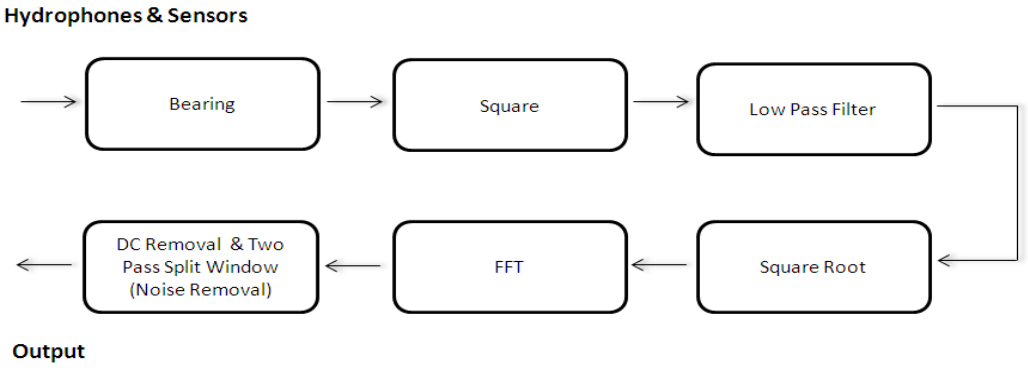

Figure 3.4: DEMON - Low Pass Filter

3.2.2 Low Frequency Analysis and Ranging (LOFAR)

Low Frequency Analysis and Ranging (LOFAR) [22] is a broadband signal analysis technique and furnishes machinery characteristics of the object i.e ships and vessels. Moreover, it provides detailed knowledge of machinery noise of the target to the sonar operator. Figure 3.5 depicts block diagram implementing LOFAR.

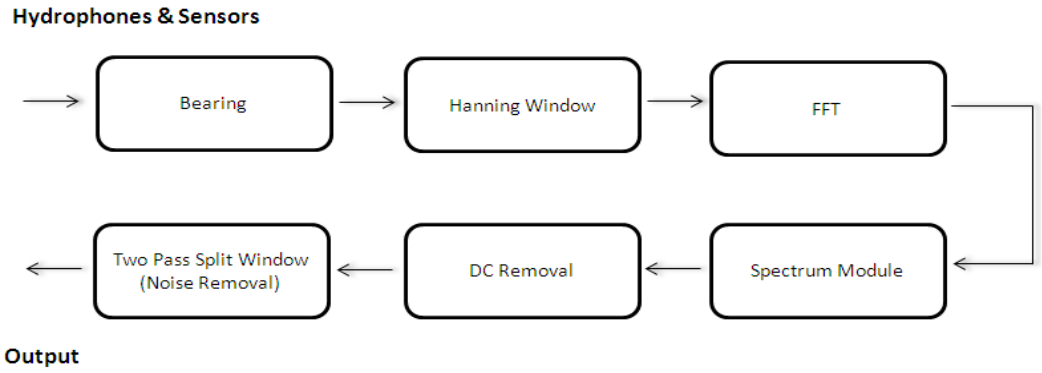

Figure 3.5: Low Frequency Analysis & Ranging (LOFAR)

On estimation of direction of arrival (DOA), captured signal is chopped into pieces

of short duration referred to as frames. Later, ensembles are passed through a window, Hann window, to select the range of interest. Filtered ensembles are frequency transformed using Fourier transform (FFT). Then, DC component is removed from the resulting spectrum and normalized using TPSW method to remove the presence of any bias due to background noise. Implementation of LOFAR is usually made using two schemes. Their brief details are in the following.

Bartlett

Bartlett method is used for spectrum estimation. It gives good frequency resolution. In Bartlett method frames are not overlapped, no-overlapping gives better frequency resolution but could result in loss of information because windowing function tends to suppress the energy of the samples near the boundaries of the frame segment to avoid spectral leakage. Having said that, it won't give a fair reflection of the original signal in the spectrum estimate.

Welch

As already discussed, LOFAR is the estimation of spectrum. Welch method [82] is used to estimate the power spectral density of the signal. It is an improved form of the standard peridiogram estimation techniques. It avoids noisy components getting into the spectra in exchange for good frequency resolution. Overlapping between two consecutive frames is 50%. Further, Welch method gives better spectrum estimates as compared to standard spectrum estimation techniques.

3.2.3 Linear Predictive Analysis (LPA)

Foundations of linear predictive analysis (LPA) [46] are built on adaptive differential pulse code modulation (ADPCM). In linear predictive analysis, prediction of $x[n]$ is done based on past samples i.e. $x[n-1], x[n-2], \ldots\ldots, x[k]$. Goal is to reduce the overall prediction error. Equation 3.2.3.1 depicts the objective function for linear predictive analysis. Main objective is to minimize error to make more accurate predictions. Linear

predictive analysis is quite widely used for acoustic signal analysis and prediction. It has most common applications in speech signal processing where it is used to mimic the speech generation model.

$$e[n] = x[n] - \sum_{k=1}^{P} \alpha_k x[n-k] \tag{3.2.3.1}$$

Figure 3.6 shows system model of ADPCM, which quite closely relates to linear predictive analysis whereas figure 3.7 illustrates system diagram implementing linear predictive analysis.

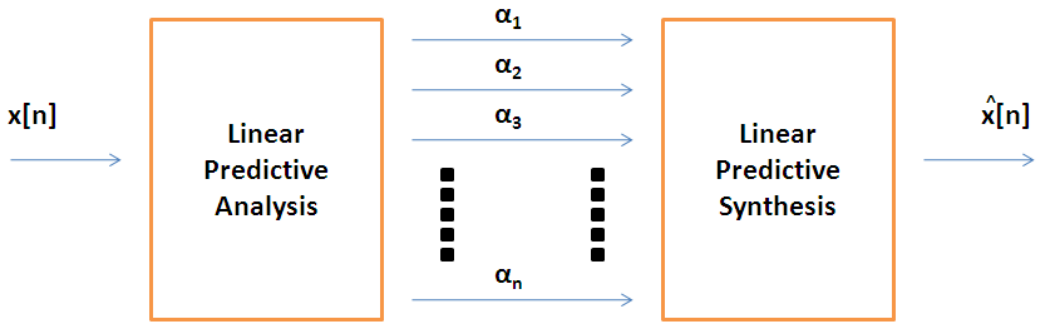

Figure 3.6: Adaptive Linear Pulse Code Modulation (ADPCM)

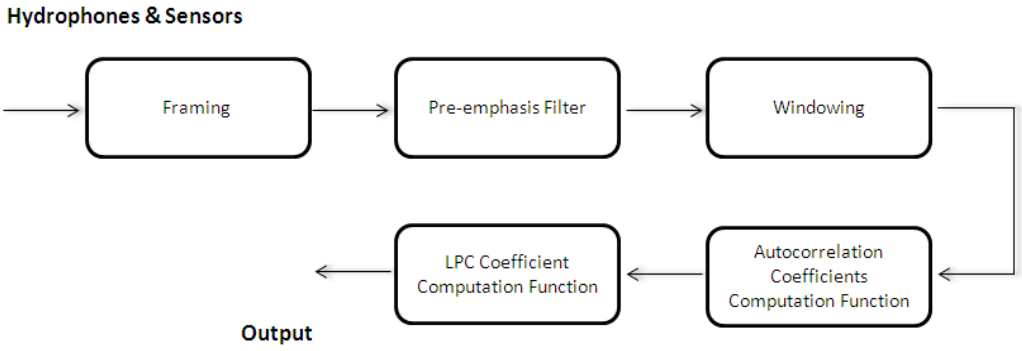

Figure 3.7: Linear Predictive Analysis (LPA)

Error Equation

Modified and frequency transformed version of equation 3.2.3.1 is given in equation 3.2.3.2 below,

$$X[z] = \left[\frac{1}{1 - \sum_{k=1}^{P} \alpha_k z^{-k}} \right] .E[z] \tag{3.2.3.2}$$

Speech Model

In speech or acoustic signal generation, the vocal tract governs the nature of the sound produced. There are several speech production models present to date. Amongst them, Formant or Vocal Tract model has been the most renowned one. Further, speech or acoustic models are important because if we can manage to make an estimate of the acoustic signal generating sources then we can predict the characteristics of speech or acoustic source with near accuracy.

Formant/ Vocal Tract Model

Formants are the fundamental frequencies at which the vocal tract resonates. Moreover, it is the concentration of acoustic energy around a particular frequency in the acoustic sequence. There are several formant bands at different positions on the frequency-scale, roughly at 1000 Hz intervals. Each formant corresponds to a frequency or a resonance in the vocal tract.

$$H[z] = \frac{1}{1 - b_1 z^{-1} + b_2 z^{-2}} \tag{3.2.3.3}$$

According to formant speech model, input to the vocal tract are the excitation pulses i.e. Impulses. Vocal tract is followed by a mouth cavity. Both the vocal tract and the mouth-cavity can be modelled using a digital filter having one-pole. Thus, a formant would be needing a two-pole filter to mimic the vocal tract resonating at a particular frequency or formant. Equation 3.2.3.3 depicts the transfer function of the speech model for a formant. It is just the cascading of two one-pole filters. Further, there is an existence of 4 to 5 formants in the human speech bandwidth (up to 3 KHz) and an additional

filter is also needed for spectral compensation as radiation is a frequency dependent term and it is done using a one-pole filter. Having said that, 4 to 5 formants exist in the range of human speech bandwidth with each formant needing a two-pole filter while a one-pole filter is needed for spectral compensation. Therefore, a 9 to 11 order filter will be needed to model the vocal tract for bandwidth of around 4 KHz. This means, we will be needing an odd-order filter. Overall transfer function of the filter representing vocal tract in ranges of human speech bandwidth is given by equation 3.2.3.4,

$$H[z] = \frac{1}{1 - c_1 z^{-1} - c_2 z^{-2} - c_3 z^{-3} .... - c_q z^{-q}} \tag{3.2.3.4}$$

where, q is the order of the filter and it will be $N + 1$. Equation 3.2.3.4 is rewritten as equation 3.2.3.5,

$$H[z] = \frac{1}{1 - \sum_{k=1}^{q} c_k z^{-k}} \tag{3.2.3.5}$$

According to the speech model, input to the vocal tract are the excitation pulses, so the transfer function can be as expressed in equation 3.2.3.6,

$$X[z] = \left[\frac{1}{1 - \sum_{k=1}^{q} c_k z^{-k}} \right] .I[z] \tag{3.2.3.6}$$

Here, c_k describes the position and bandwidth of the formant's resonance. The speech model in figure 3.8 looks similar to ADPCM model discussed earlier and illustrated as in figure 3.6.

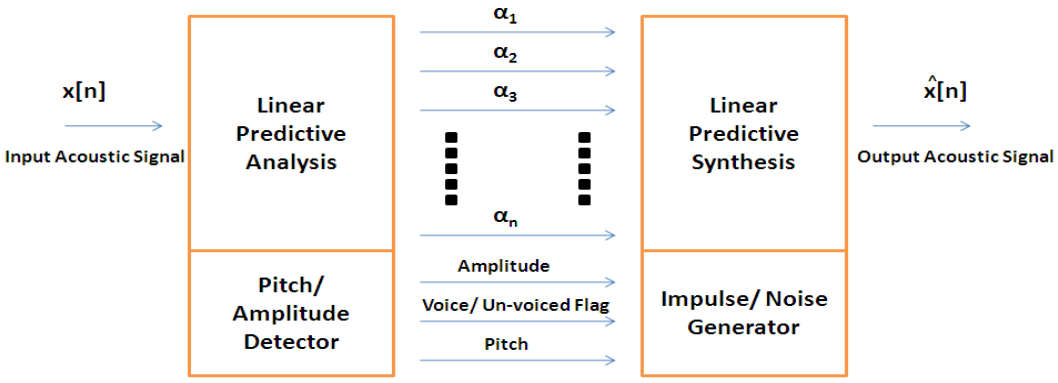

Figure 3.8: Low-Bit Rate Speech Coder/Decoder

Process in fig. 3.8 can be used to model any acoustic signalling source. An estimate of α's will be required to model an acoustic source. We need the first half of the block

presented in figure 3.8 to build an acoustic model. On calculation of α's, denominator in equation 3.2.3.6 needs to be factorized to calculate formant frequencies. Calculations involving formant frequencies are computationally very demanding and expensive. Moreover, values of α's will be obtained on minimizing the error i.e. Mean Square Error (MSE). Following represents the mathematical equation of mean square error,

$$M = \sum_{\forall n} e^2(n) = \sum_{\forall n} \left[x[n] - \sum_{k=1}^{P} \alpha_k x[n-k] \right]^2 \tag{3.2.3.7}$$

For calculating 'α', mean square error should be to global minimum. Mathematically, it is given by,

$$\frac{dm}{d\alpha_j} = 0 \tag{3.2.3.8}$$

Differentiating equation 3.2.3.7 with respect to 'α' and equating it to 0,

$$-2 \sum_{\forall n} x[n-j] \left[x[n] - \sum_{k=1}^{p} \alpha_k x[n-k] \right] = 0$$

$$\sum_{k=1}^{p} \alpha_k \sum_{\forall n} x[n-j]x[n-k] = \sum_{\forall n} x[n]x[n-j]$$

$$where,\ 1 \leq j \leq P \tag{3.2.3.9}$$

Hence from equation 3.2.3.9, it can be inferred that for calculating 'p' unknowns, p-equations are to be solved. Further, the problem is computationally very demanding and involves calculation of inverse of a $p \times p$ matrix. So, a lesser complicated procedure for calculating the 'p' unknowns is to be used.

Computational Aspects

Further, equations discussed have been for a sequence of length n, now we want to calculate it for a segment or a frame. Modifying previous equations for a finite length sequence, $s_n(m)$. Mathematically, it is given by,

$$\sum_{k=1}^{p} \alpha_k \sum_{m} s_n[m-i]s_n[n-k] = \sum_{m} s_n[m]s_n[m-i]$$

$$where,\ 1 \leq i \leq P \tag{3.2.3.10}$$

Solving above p equations will yield p-α coefficients. Looking at R.H.S of the above equation, it looks like an autocorrelation function, so equation 3.2.3.10 can be rewritten as,

$$\phi_n(i,k) = \sum_m s_n(m-i)s_n(m-k) \tag{3.2.3.11}$$

Putting values from equation 3.2.3.11 in equation 3.2.3.10,

$$\sum_{k=1}^{p} \alpha_k \phi_n(i,k) = \phi_n(i,0)$$
$$where, i = 1,2,3,....,p \tag{3.2.3.12}$$

In terms of autocorrelation, mean square error (MSE) is given by,

$$\begin{aligned} E_n &= \sum_m s_n^2(m) - \sum_{k=1}^{p} \alpha_k \sum s_n(m)s_n(m-k) \\ &= \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k) \end{aligned} \tag{3.2.3.13}$$

There are several approaches available in literature to solve for the 'p' α-coefficients. i.e. Autocorrelation Method, Lattice Method etc. In this study, Autocorrelation Method [58] has been used to calculate the α-coefficients.

Autocorrelation Method

Error Equation,

$$E_n = \sum_{m=0}^{N+P-1} e_n^2(m) \tag{3.2.3.14}$$

Autocorrelation method will be applied to a segment or frame instead of all n samples. Mathematically,

$$s_n[m] = s[m+n]w[m]$$
$$where,\ 0 \leq m \leq N-1 \tag{3.2.3.15}$$

where, $s_n(m)$ outside the interval will be considered zero, that is, $0 \leq m \leq N-1$.

The autocorrelation function can be written as,

$$\phi_n(i,k) = \sum_{m=0}^{N+P-1} s_n(m-i)s_n(m-k)$$

$$where, \qquad 1 \leq i \leq p$$

$$0 \leq k \leq p \qquad (3.2.3.16)$$

cosmetic changes in equation 3.2.3.16 will shape the equation as,

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(n)s_n(m+i-k)$$

$$= R_n(i-k) \qquad (3.2.3.17)$$

where,

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m)s_n(m+k)$$

and

$$\phi(i,k) = R_n(|i-k|)$$

For finding α's, equation 3.2.3.12 needs to be solved, re-writing equation in terms of R_n,

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i)$$

$$where,\ 1 \leq i \leq p \qquad (3.2.3.18)$$

The p simultaneous equations can be written in form of a matrix as given in equation below,

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(2) & R_n(3) & \dots & R_n(p-2) \\ R_n(2) & R_n(3) & R_n(4) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \vdots \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \vdots \\ \vdots \\ \vdots \\ R_n(p) \end{bmatrix}$$

(3.2.3.19)

Solving equation 3.2.3.19 for linear predictive coefficients involves expensive mathematical calculations i.e. inverse of a $p \times p$ matrix, where p are the number of linear predictive coefficients. If looked closely at equation 3.2.3.19, matrix on R.H.S looks a special type of matrix known as **Toeplitz Matrix**.

There are several numerical approaches available to solve this kind of a matrix i.e. Durbin's recursive procedure [58] and Bareiss Method [7] etc. Set of equations describing Durbin's recursive method are discussed below.

Durbin's Recursive Procedure

1. $E^{(0)} = R(0)$
2. $k_i = \left[\frac{R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)}{E^{(i-1)}} \right]$ $\quad where,\ 1 \leq i \leq p$
3. $\alpha_i^{(1)} = k_i$
4. $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$ $\quad where,\ 1 \leq j \leq i - p$
5. $E^{(i)} = (1 - k_i^2) E^{(i-1)}$

Equations (2) to (5) are solved recursively for, $i = 1, 2, 3, \ldots, p$, coefficients. As 'i' reach p^{th} iteration, linear predictive coefficients can be represented by a set as under,

$$\alpha_j = \alpha_j^{(p)} \qquad where, 1 \leq j \leq p \tag{3.2.3.20}$$

Coefficients in equation 3.2.3.20 are said to be the linear predictive coefficients and will be referred to as the feature vector representing characteristics of the acoustic source.

3.2.4 Linear Predictive Cepstral Coefficient (LPCC)

Linear Predictive Cepstral Coefficient (LPCC) [58] is a commonly used technique in acoustic and speech signal processing. The concept behind LPCC is similar to LPA, that is, modelling of the human vocal tract using an all-pole digital filter. Though, it adds one more step compared to LPA, that is, the cepstral analysis. Advantage of taking cepstrum is that it further refines the coefficients extracted using LPA. Figure 3.9 represents the overall system model for LPCC.

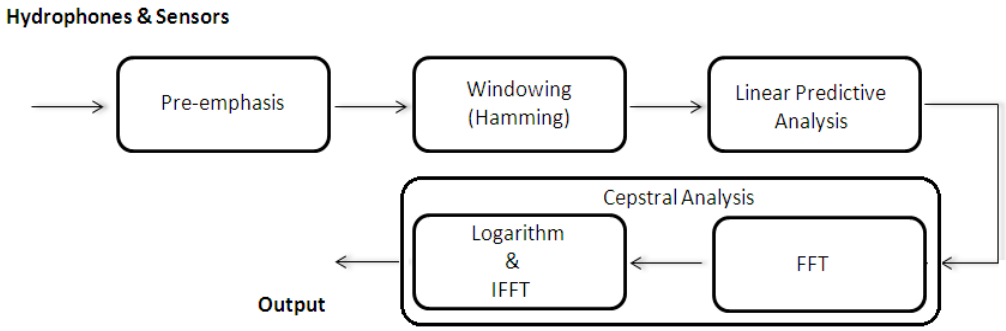

Figure 3.9: Linear Predictive Cepstral Coefficient (LPCC)

Pre-emphasis and Hamming Window

Processing of acoustic sequence starts with pre-emphasis [59], which is applied to the complete signal. It is modelled using a first-order low pass digital filter. The idea behind pre-emphasis is to spectrally flatten the signal. Equation below represents transfer function of the said filter,

$$H_p(z) = 1 - az^{-1} \tag{3.2.4.1}$$

where, 'a' is a constant and it's typical value is taken as 0.97.

After pre-emphasis, the signal is broken into segments referred as frames. Duration of each frame should be kept between $15ms$ to $25ms$, duration more than that would have adverse effects on the detection process. After framing, each frame is windowed i.e. it will be multiplied with a windowing function. Idea behind windowing is to mitigate spectral leakage because framing an aperiodic or random signal can lead to discontinuities

yielding high frequency components which aren't part of the actual signal. Windowing function used is hamming and it is given by,

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & otherwise \end{cases} \tag{3.2.4.2}$$

Mathematically, windowing operation can be expressed with the following equation,

$$s_m[n] = s[n]w[n-m] \tag{3.2.4.3}$$

where, $s[n]$ represents the complete signal, $s_m[n]$ represents windowed frame of a signal recorded at instant m, $w[n]$ represents the Hamming window function and 'N' represents the length of the Hamming window. Moreover, 'm' can be defined as the step size or the time shift of the windowing function. Frames can be obtained by moving the windowing function by 'm' steps in time. To avoid any loss of information, overlapping between two consecutive frames has been kept to 12.5 ms in this study.

As discussed earlier, windowing function shape is of key importance. Uniform window is not suggested in any case even if we don't have any information about the characteristics of the signal. Rectangular window can cause severe spectral leakage. Other types of windowing functions also exist but their usage is subject to the characteristics of the signal. After windowing, the signal is passed on to the next stage.

Linear Predictive Analysis (LPA)

Linear predictive analysis will be applied to each windowed segment to obtain linear predictive coefficients. Durbin's Recursive Method, a numerical approach for implementing autocorrelation method will be used for calculating linear predictive coefficients. The process will be similar as discussed in section 3.2.3.

Cepstral Analysis

Cepstral analysis [58] is a process for finding cepstrum of a signal. Cepstrum refers to as "the inverse frequency transformation of log of a signal's power spectrum". The word cepstrum is formed by moving around the letters of spectrum. It refers to as the

time-domain representation of the signal. However, quefrency, a special term is used instead of time to describe basis of the cepstrum. It is formed by shuffling the letters of the term frequency. Following equation represents the cepstrum of a signal.

$$\hat{s}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} ln[S(\omega)] e^{j\omega n} d\omega \tag{3.2.4.4}$$

where, $S(\omega)$ represents power spectrum of a windowed frame sequence and $s[n]$ represents the resulting cepstrum.

The idea behind cepstral analysis is de-convolution. It is quite frequently used in speech signal processing. Cepstral analysis is primarily used to extract the characteristics of the vocal tract from the acoustic spectrum. According to formant speech model, an acoustic signal is the result of convolution of the vocal tract and the glottal excitation pulses. Mathematically, speech generation model can be expressed with the following equation.

$$s[n] = v[n] * u[n] \tag{3.2.4.5}$$

where, $v[n]$ represents behaviour of the vocal tract, $u[n]$ represents glottal excitation pulses. It is a quasi-periodic sequence of impulses for voiced part of speech. Mathematically, above can be expressed in frequency domain as,

$$S(\omega) = V(\omega)U(\omega) \tag{3.2.4.6}$$

With reference to equation 3.2.4.4, application of logarithm splits the product of the two spectrum inside the integral into a linear summation. Moreover, resulting sequence will be the sum of the vocal tract cepstrum and the glottal excitation cepstrum.

$$\hat{s}[n] = \hat{v}[n] + \hat{u}[n] \tag{3.2.4.7}$$

where, $s[n]$ is the acoustic sequence, $v[n]$ represents behaviour of the vocal tract and $u[n]$ represents the glottal excitation pulses.

Cepstrum hold two distinct properties that makes it useful and they are listed below:

1. Cepstrum of a periodic signal is always periodic.

2. Cepstrum of a system with random behaviour is always decaying, which tends to zero as n, quefrency, approaches infinity.

Since, $v[n]$ represents the impulse response of the vocal tract filter, $u[n]$ is a quasi-periodic impulse sequence and from the above two properties it can be inferred that low-frequency part of the cepstrum represents the vocal tract while the high frequency content represents the glottal excitation pulses, so first few coefficients of the resulting cepstrum has all the information about the vocal tract and the lateral half represents glottal excitations. Therefore, only first few coefficients of the cepstrum are kept to form the feature vector. These are termed as cepstral coefficients. Further, cepstral analysis is performed on the LPA coefficients resulting in linear predictive cepstral coefficients.

3.2.5 Perceptual Linear Prediction (PLP)

Perceptual Linear Prediction (PLP) [36] [34] [35], also known with the name of Bark Frequency Cepstral Coefficient (BFCC) (due to the use of a special filter bank, known as the Bark filter bank), is one of the most renowned feature extraction techniques exist in the domain of speech signal processing. It models the human auditory system. The principle of PLP lies within three major characteristics of the auditory nerve and they are as follows:

1. Critical Band Frequency Selectivity

2. Equal-loudness curve

3. Intensity-Loudness Power Law

Initial step is to calculate spectrum of a frame obtained after windowing. Then, resulting spectrum undergoes a special type of filter known as the Bark filter. It implements the critical band selectivity property of the auditory nerve. After that, the outgoing filter coefficients are transformed as per the equal-loudness curve function that emulates the human hearing sensitivity. Next step is to compress the weighted filter coefficients using intensity-loudness power law, which represents the relationship between perceived loudness and signal intensity. Inverse Fourier transform is taken of the resulting compressed coefficients and then, linear predictive analysis technique is applied to the transformed sequence. Final step of the process is to perform cepstral analysis, which can be done using equation 3.2.4.4 or alternate recursive procedures can be used. Figure 3.10 represents the overall system model for implementation of PLP.

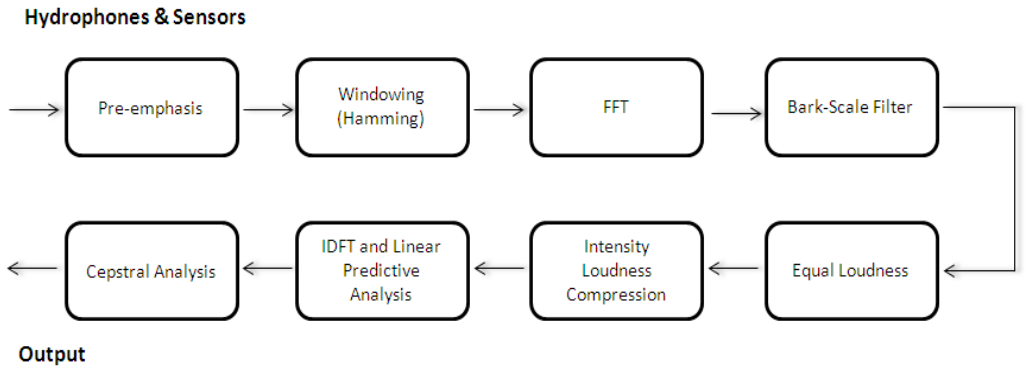

Figure 3.10: Perceptual Linear Prediction (PLP)

Steps for implementing Perceptual Linear Prediction (PLP) are discussed in the following,

Hamming Window and FFT

First step is to segment the acoustic signal and apply windowing function to each frame. Hamming window is the windowing function used in this case. Purpose and details of windowing have already been discussed in section 3.1. Next step is to take Fourier transform (FFT) of the windowed sequence. Then, resulting spectrum, $S(\omega)$, will be passed to the next stage for further processing.

Bark Scale Filter

In PLP, a filter bank is used based upon a non-linear frequency-scale referred to as the bark-scale. Mathematical relationship between linear frequency-scale and bark-scale can be expressed with the following equation.

$$f_{Bark} = 6\ln(\frac{f}{600} + ((\frac{f}{600})^2 + 1)^{0.5}) \tag{3.2.5.1}$$

where, f represents frequency in hertz and f_{Bark} represents frequency on bark-scale in barks.

Moreover, filters are evenly spread along the bark-scale, though, spacing is with respect to their center frequencies. Moreover, it was suggested in an article [34] that filters along bark-scale should approximately be 1 *bark* apart. Also, that position of first and last filter on the bark-scale should be 0 *bark* and Nyquist frequency (highest frequency component present in the signal) *bark*, respectively. Following equation depicts the mathematical model of the bark filter bank.

$$\psi = \begin{cases} 0 & for,\ f_{Bark} - fc(Bark) < -2.5 \\ 10^{f_{Bark} - f_{c(Bark)} + 0.5} & for,\ -2.5 \leq f_{Bark} - fc(Bark) \leq -0.5 \\ 1 & for,\ -0.5 \leq f_{Bark} - fc(Bark) \leq 0.5 \\ 10^{-2.5(f_{Bark} - f_{c(Bark)} - 0.5)} & for,\ 0.5 \leq f_{Bark} - fc(Bark) \leq 1.5 \\ 0 & for,\ f_{Bark} - fc(Bark) > 1.3 \end{cases} \tag{3.2.5.2}$$

where, '$f_{c(Bark)}$' is the center frequency of a filter in the bark filter bank, 'ψ' represents weight of frequencies along bark-scale. Figure 3.11 gives an idea of filter placement and position in the filter bank along the linear frequency-scale. It can be seen that all filters in the filter bank are non-linearly positioned on the linear frequency-scale. Also, that filters positioned high up the linear frequency scale have wider bandwidth compared to the ones positioned at low frequency levels. This describes the non-linear relationship between the bark-scale and the linear frequency-scale. However, shapes of the filters are identical in the Bark filter bank.

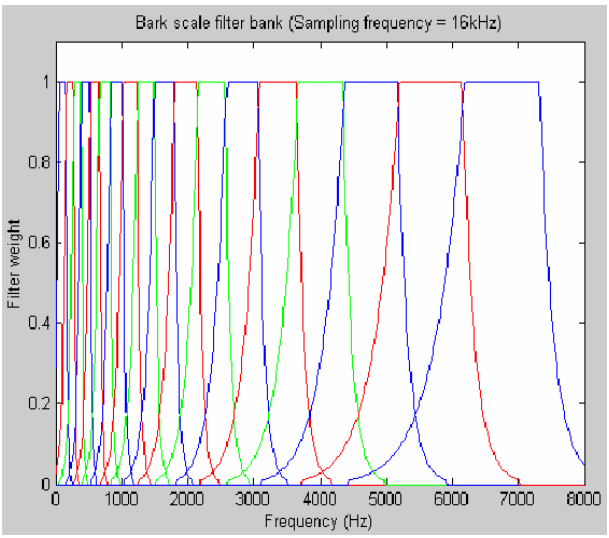

Figure 3.11: Bark-scale Filter Bank [52]

Moreover, output of each filter is represented as the sum of the product of FFT

spectrum and the filter's weights. Mathematically, it can be expressed as,

$$X_m = \sum_{k=0}^{N/2-1} |S[k]|^2 |\psi_m[k]|$$
$$where,\ 2 \leq m \leq M-1 \tag{3.2.5.3}$$

where, X_m is output of the m^{th} filter in the filter bank, $|S[k]|^2$ represents power spectrum of the windowed sequence, it is of length N and $|\psi_m[k]|$ represents the magnitude response of the m^{th} filter in the filter bank on linear frequency-scale. It is also of note that output for first and last filter in the filter bank are not evaluated in this step. It is because that these two filters have almost identical shapes as of their adjacent filters. Output for these two filters will be calculated in the next step.

Equal-Loudness Curve

Equal-loudness curve [34] function models human hearing sensitivity with respect to frequency levels. It has two basic mathematical models catering for different frequency ranges. One for applications whose Nyquist frequency is 5 KHz or below and the other for applications where Nyquist frequency is above 5 KHz. Equations 3.2.5.4 and 3.2.5.5 represents the mathematical model for both the criterion,

$$E = \frac{(\omega^2 + 56.8 \times 10^6)\omega^4}{(\omega^2 + 6.3 \times 10^6)^2(\omega^2 + 0.38 \times 10^9)} \tag{3.2.5.4}$$

and,

$$E = \frac{(\omega^2 + 56.8 \times 10^6)\omega^4}{(\omega^2 + 6.3 \times 10^6)^2(\omega^2 + 0.38 \times 10^9)(\omega^6 + 9.58 \times 10^26)} \tag{3.2.5.5}$$

where, ω represents angular frequency in radians.

From the above two equations, equal-loudness weight for outgoing filter coefficients can be calculated by replacing the respective value of center frequency (radians) in one of the aforementioned equations, choice of equation will depend upon the Nyquist frequency.

Each outgoing filter coefficient from the filter bank will be weighted using the equal-loudness function. Mathematically, it can be expressed as,

$$X_{m(e)} = E_m X_m$$
$$where,\ 2 \leq m \leq M-1 \tag{3.2.5.6}$$

where, $X_{m(e)}$ represents output from the m^{th} filter after multiplication with equal-loudness weight and E_m represents the equal-loudness weight of the m^{th} filter.

As discussed earlier, the shape of first and last filters in the bark filter bank is identical to the adjacent filter. So, weight of the second filter after equal-loudness curve function will be assigned to the first filter and likewise, weight of next-to-last filter will be assigned to the last filter. It can be mathematically expressed as

$$X_{1(e)} = X_{2(e)}$$
$$X_{M-1(e)} = X_{M(e)}$$

Intensity-Loudness Curve

Intensity-loudness curve [34] represents human perceptibility of loudness given signal intensity at different frequencies; it models the relationship between perceived loudness and signal intensity which is non-linear in nature. In PLP, this relationship is mathematically described with cubic root. Following equation describes the intensity-loudness relationship,

$$\phi_m = (X_{m(e)})^{0.33} \quad where,\ 1 \leq m \leq M \tag{3.2.5.7}$$

where, ϕ_m represents output of a filter after intensity-loudness operation.

IDFT and Linear Predictive Analysis

Next step is to take inverse Fourier transform of the resulting coefficients after intensity-loudness operation. Then, linear predictive analysis is applied to the transformed coefficients. Further, inverse Fourier transform will be applied to the filtered coefficients that will yield autocorrelation function coefficients and LPA will be applied to the extracted autocorrelation coefficients. Notice that ϕ_m only represents one half of the spectrum i.e. till Nyquist frequency. Before applying IDFT, spectrum will be appended with its mirror to form a double-sided spectrum. Note that first and last filter weights were same as their neighbours. So, they will be excluded from the mirroring process. Mathematically,

the resulting spectrum can be expressed as in the following equation.

$$\Phi = [\phi_1 \ \phi_2 \ \phi_3 \ \ldots\ldots \ \phi_{M-1} \ \phi_M \ \phi_{M+1} \ \ldots\ldots \ \phi_3 \ \phi_2] \tag{3.2.5.8}$$

IDFT is applied to the vector shown above, where first p coefficients will be regarded as the autocorrelation function sequence, $R[n]$, for $1 \leq n \leq p$. Autocorrelation function sequence will be used as an input to perform linear predictive analysis. As discussed earlier, Durbins recursive algorithm will be used for calculating linear predictive coefficients.

Cepstral Analysis

Cepstral analysis is applied to the extracted linear predictive coefficients. The process of cepstral analysis will be same as discussed in section 3.2.4. The order of PLP coefficients will be same as that of linear predictive coefficients. There will be 'p' number of PLP coefficients calculated and will be regarded as feature vector for that particular frame sequence.

3.2.6 Mel-Frequency Cepstral Coefficient (MFCC)

Mel-Frequency Cepstral Coefficient [18] is one of the most used front-ends in speech signal processing. It is a FFT based method i.e. feature vectors are extracted from spectrum of the windowed frame. Figure 3.12 illustrates the overall feature extraction process for mel-frequency cepstral coefficient technique.

Pre-emphasis, Windowing and FFT

Like LPCC, complete acoustic signal is fed to a one-pole digital FIR filter, implementing pre-emphasis function. The concept behind pre-emphasis is to magnify the high frequency contents in the signal before the signal is processed further. The transfer function of the filter will be same as mentioned earlier. After pre-emphasis, signal will be broken into segments of duration $15ms$ to $25ms$ i.e. to avoid non-stationarity. Next, framed sequences are passed through a windowing function, namely, Hamming window i.e. to avoid spectral leakage. The impulse response of the windowing function will be same as described earlier. The windowed frame is passed onto the next stage to extract

frequency contents in the signal, that is, FFT of the windowed frame is calculated to extract spectrum of the signal, $S[k]$, for $0 \leq k \leq N-1$. Spectrum of the windowed frame is fed to the next stage for further processing.

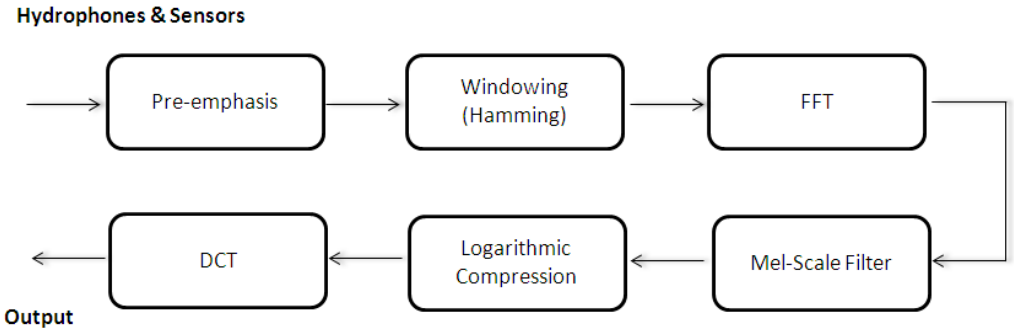

Figure 3.12: Mel-Frequency Cepstral Coefficient (MFCC)

Mel-Filter Bank

Mel-filter bank is represented with a series of band pass filters that are triangular in shape. It tries to emulate the human auditory nerve. Moreover, mel-filter is based on a non-linear frequency-scale, referred to as the mel-scale. Mel-scale represents the measure of pitches observed by humans. In a study [76], it was validated that a tone of 1000 Hz with an intensity of about 40 dB or above listeners threshold, is regarded as having a pitch of 1000 $mels$. Mel-scale is almost equal to the linear frequency-scale below 1000 Hz. Further, listener's perceived pitch increments with longer frequency intervals above $1000Hz$ reference point and proves the relationship between mel-scale and linear frequency-scale to be non-linear. Moreover, the relationship is logarithmic above the 1000 Hz reference point. Relationship between linear frequency-scale and mel-scale is given by the following equation,

$$f_{mel} = 1127.01 \times log(f/100 + 1) \tag{3.2.6.1}$$

Moreover, inverse relationship can be mathematically expressed as in equation given below,

$$f = 700(e^{\frac{f_{mel}}{1125}} - 1) \tag{3.2.6.2}$$

where, f_{mel} represents frequency on mel-scale and f represents frequency on linear-scale.

As described earlier, filters in mel-filter bank are triangular bandpass filters. These filters overlap each other in such a way that boundary of one filter lies at the center frequency of its adjacent filter i.e. lower boundary of the filter is positioned at the center frequency of the preceding filter and upper boundary of the filter is positioned at center frequency of the next filter. Moreover, maximum magnitude response of a filter lies at its center frequency, that is, the top vertex of a triangle. Further, magnitude response of all filters in the filter bank is normalized to unity. Following equation represents the mathematical model of a mel-filter bank.

$$H_m[k] = \begin{cases} 0 & for,\ k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & for,\ f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & for,\ f(m) \leq k \leq f(m+1) \\ 0 & for,\ k > f(m+1) \end{cases} \tag{3.2.6.3}$$

Filters in the mel-filter bank are evenly spread along the mel-scale. Center frequencies of filters in the mel-filter bank are evaluated using the following equation. Where, 'm' represents the m^{th} filter in the filter bank,

$$f_{cm}(mel) = f_l(mel) + \frac{m(f_h(mel) - f_l(mel))}{M+1} \qquad where, 1 \leq m \leq M \tag{3.2.6.4}$$

where, $f_{c(mel)}$ represents center frequency of a filter on mel-scale in mels. $F_L(mel)$ and $F_H(mel)$ are the lower and upper bounds of the mel-frequency scale. There are 'M' filters lying in between this range. Following figure shows magnitude response of the filter bank along linear frequency-scale.

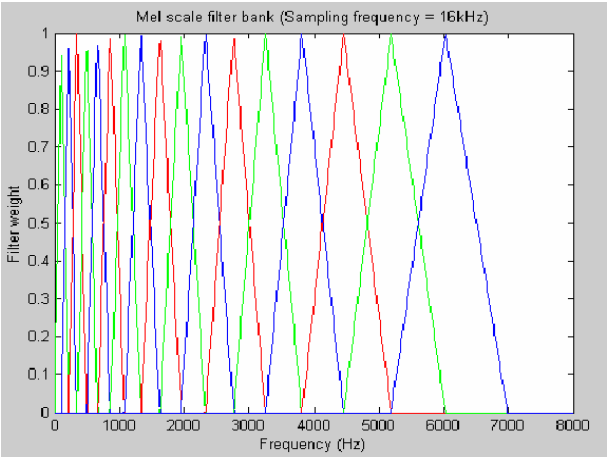

Figure 3.13: Mel Filter Bank [52]

It is of note that all filters are identical in shape and as we move along the linear frequency-scale, bandwidth of these triangular bandpass filters gets wider and the center frequencies of the respective filters are not evenly spaced. This goes to show that relationship between mel-scale and linear frequency-scale is non-linear.

Moreover, if $|H_m[K]|$ represents magnitude response of the m^{th} filter, where k is the frequency domain index, then output, X, of the filter can be expressed by the following equation,

$$X_m = \sum_{k=0}^{N/2} |S[k]|^2 |H[k]|$$
$$where, 1 \leq m \leq M \qquad (3.2.6.5)$$

where, $S[k]$ represents the spectrum of the windowed frame. Note that in equation 3.2.6.5, summation is taken till $N/2$ points because lateral half is mirror of the first half of the spectrum.

Logarithmic Compression

To model human perceived loudness with respect to signal intensity, logarithmic compression is applied. Mathematically, output after logarithmic compression can be expressed with the following equation,

$$X_m(\ln) = \ln(X_m)$$
$$where, 1 \leq m \leq M \qquad (3.2.6.6)$$

where, $X_{m(ln)}$ represents the output of the m^{th} filter after logarithmic compression function.

Discrete Cosine Transform (DCT)

After logarithmic compression, resulting coefficients are passed to the next stage for the concluding final step of cepstral analysis i.e. Discrete Cosine Transform (DCT). It is applied to the filter coefficients to de-correlate them. Only first few coefficients are kept after the application of DCT as they completely defines the vocal tract and are regarded

as the cepstral coefficients. Mathematically, the k^{th} MFCC coefficient can be expressed by equation 3.2.6.7.

$$MFCC_k = \sqrt{2/M} \sum_{m=1}^{M} \cos(\frac{\pi k(m-0.5)}{M}) \quad where, 1 \leq k \leq P \tag{3.2.6.7}$$

Suppose, 'p' is the order of the mel-filter bank. Then, cepstrum will have 'p' coefficients after DCT. Only first few coefficients are kept to form the feature vector.

3.2.7 Gammatone Cepstral Coefficient (GTCC)

Gammatone Cepstral Coefficient [2] is another FFT-based front-end used in speech signal processing. It also tries to mimic human auditory system. It is relatively a newer technique compared to other conventional cepstral analysis methods and works well in noisy conditions compared to other acoustic signal detection front-ends [52]. This technique is based upon a gammatone-filter bank, that mimics the frequency selectivity property of the human auditory system with a series of bandpass filters. Like MFCC, feature vector is calculated using spectral content of the windowed frame. Following figure illustrates the overall feature extraction process of GTCC.

Steps of feature extraction are discussed in the following.

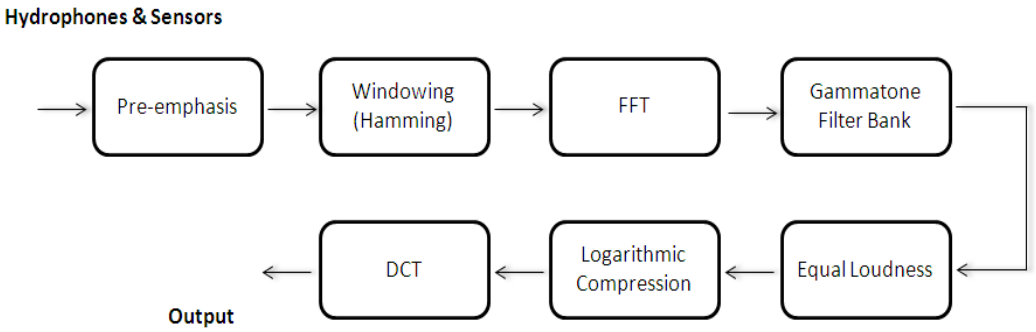

Figure 3.14: Gammatone Cepstral Coefficient (GTCC)

Hamming Window and FFT

First step of the algorithm is similar to other cepstral analysis techniques, that is, dividing the signal into small segments regarded as frames. Segments are later passed through a windowing function i.e. Hamming window. After windowing, windowed-frame is frequency transformed using fast Fourier transform (FFT). The N-point FFT represents the spectrum of the signal, that is, $S[k]$ for $0 \leq k \leq N-1$.

Gammatone Filterbank

Gammatone filter bank comprised of a series of bandpass filters, which represents frequency selectivity property of the auditory system. Response of a filter in the filter bank explained in [73] and can be mathematically expressed as in the equation below,

$$g(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \phi) \tag{3.2.7.1}$$

where, 'a' is a constant. Usually its value is taken as 1. Moreover, 'ϕ' represents the phase shift, 'f' represents the frequency and 'b' represents the bandwidth of the filter in Hz.

According to a research [73], center frequency and bandwidth of each filter can be calculated using the equivalent rectangular bandwidth (ERB). The ERB of a filter is related to its center frequency. Concept behind ERB is that human cochlea can be modelled using a series of rectangular bandpass filters; the bandwidth of each filter is termed as equivalent rectangular bandwidth (ERB). Following equation describes the mathematical relationship between auditory filter's ERB and center frequency [29].

$$ERB(f_c) = 24.7(\frac{4.37 \times fc}{1000} + 1) \tag{3.2.7.2}$$

Also, Slaney [73] suggested that bandwidth of a gammatone filter should approximately be 1.019 times the ERB from center frequency of the said filter. It can be mathematically expressed as,

$$b = 1.019\ ERB \tag{3.2.7.3}$$

It was also recommended [73] that a 4^{th} order gammatone filter should be good to model an auditory filter. This answers two unknowns; one being the order and second being the bandwidth of the filter. Now, the question is how much spaced should two

consecutive filters be in the filter bank. In an article, it was suggested that a filter should be a fraction of ERB spaced from a preceding filter. Following equation shows the relationship to calculate center frequency of a filter in the filter bank [73].

$$f_{cm} = \frac{-1000}{4.37} + (f_h + \frac{1000}{4.37})e^{\frac{m}{M}(-\ln(f_h+\frac{1000}{4.37})+\ln(f_l+\frac{1000}{4.37}))}$$
$$where,\ 1 \leq m \leq M \qquad (3.2.7.4)$$

where, f_{cm} represents center frequency of a gammatone filter. Moreover, F_l and F_h are the lower and upper bounds of the filter bank in Hz, respectively. There are total M filters in the gammatone filter bank distributed between the range F_l and F_h along the linear frequency-scale.

To implement gammatone filter bank an efficient approach was proposed; each 4^{th} gammatone filter was implemented using cascading of four filters. Each cascading stage is build using a 2^{nd} order filter. Equation 3.2.7.5 to 3.2.7.8 represents the transfer function of each of the cascading stages

$$H^1(z) = \frac{-2T + (2Te^{-2\pi bt}\cos(2\pi f_c T) + 2\sqrt{3+2^{3/2}}Te^{-2\pi bt}\sin(2\pi f_c t))z^{-1}}{-2 + 4e^{-2\pi bt}\cos(2\pi f_c T)z^{-1} - 2e^{-4\pi bt}z^{-2}} \qquad (3.2.7.5)$$

$$H^2(z) = \frac{-2T + (2Te^{-2\pi bt}\cos(2\pi f_c T) - 2\sqrt{3+2^{3/2}}Te^{-2\pi bt}\sin(2\pi f_c t))z^{-1}}{-2 + 4e^{-2\pi bt}\cos(2\pi f_c T)z^{-1} - 2e^{-4\pi bt}z^{-2}} \qquad (3.2.7.6)$$

$$H^3(z) = \frac{-2T + (2Te^{-2\pi bt}\cos(2\pi f_c T) - 2\sqrt{3-2^{3/2}}Te^{-2\pi bt}\sin(2\pi f_c t))z^{-1}}{-2 + 4e^{-2\pi bt}\cos(2\pi f_c T)z^{-1} - 2e^{-4\pi bt}z^{-2}} \qquad (3.2.7.7)$$

$$H^4(z) = \frac{-2T + (2Te^{-2\pi bt}\cos(2\pi f_c T) + 2\sqrt{3-2^{3/2}}Te^{-2\pi bt}\sin(2\pi f_c t))z^{-1}}{-2 + 4e^{-2\pi bt}\cos(2\pi f_c T)z^{-1} - 2e^{-4\pi bt}z^{-2}} \qquad (3.2.7.8)$$

where, T represents sampling interval of the system. It can drawn from the above discussion that complete response of a gammatone filter is the product of responses of the cascading filters. Mathematically, it can be expressed as,

$$H(z) = H^1(z)H^2(z)H^3(z)H^4(z) \qquad (3.2.7.9)$$

Here, we are only interested in the magnitude response of the filters, $|H(\omega)|$. Magnitude response of each filter is extracted and normalized to unity. Figure 3.15 illustrates the normalized magnitude response of the gammatone filter bank.

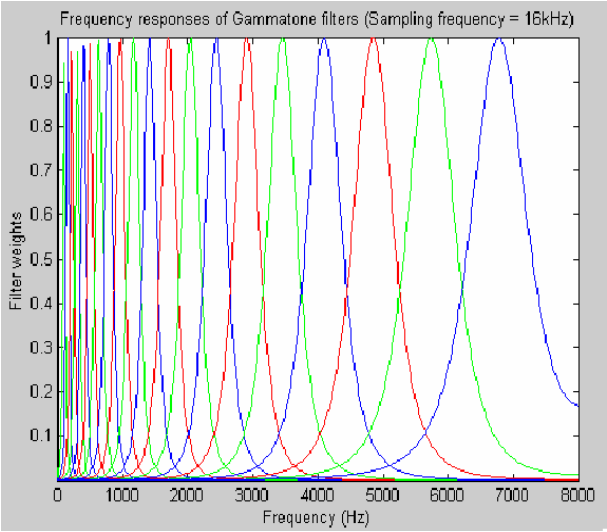

Figure 3.15: Gammatone Filter Bank [52]

As discussed previously, after windowing FFT is applied to the windowed-frame. Moreover, the resulting power spectrum, $|S[k]|^2$, is fed to the next stage, that is, to the gammatone filter bank. Further, power spectrum till Nyquist frequency of the signal is kept i.e. only half of the N-points because the lateral is the mirror of the first half. The output of a gammatone filter, X_m, can be mathematically expressed with the following equation,

$$X_m = \sum_{k=0}^{N/2-1} |S[k]|^2 |H_m[k]| \qquad where,\ 1 \leq m \leq M \tag{3.2.7.10}$$

where, N represents the number of FFT points and $H[k]$ is the frequency response of a filter in the gammatone filter bank.

Equal-Loudness

As in PLP, equal-loudness function implements the sensitivity of human hearing. Filter outputs are weighted according to their center frequencies. First and last filters in the filter bank will not be weighted as the respective filters have identical shape same as their adjacent filters. These two filter coefficients will be processed in the next step.

Output of a gammatone filter, Xm, after weighting as per the equal-loudness function can be mathematically expressed by the following equations,

$$X_{m(e)} = E_m X_m \quad where,\ 2 \leq m \leq M-1 \tag{3.2.7.11}$$

Equal-loudness function, for applications having Nyquist frequency below 5 KHz,

$$E(w) = \frac{(w^2 + 56.8 \times 10^6)w^4}{(w^2 + 6.3 \times 10^6)^2(w^2 + 0.38 \times 10^9)} \quad for,\ f_m \leq 5KHz \tag{3.2.7.12}$$

and, for applications having Nyquist frequency above 5 KHz,

$$E(w) = \frac{(w^2 + 56.8 \times 10^6)w^4}{(w^2 + 6.3 \times 10^6)(w^2 + 0.38 \times 10^9)(w^6 + 9.58 \times 10^{26})} \quad for,\ f_m > 5KHz \tag{3.2.7.13}$$

Logarithmic-Compression

Next step is to apply the logarithmic-compression function to the weighted filter coefficients. It models the perceived loudness of human auditory nerve in response to the signal intensity. Following equation represents the mathematical relationship between perceived loudness and signal intensity,

$$X_{m(\ln+e)} = \ln(X_{m(e)}) \quad where,\ 1 \leq m \leq M \tag{3.2.7.14}$$

As discussed, first and last filters in the filter bank will be assigned weights same as of their adjacent filters. Mathematically, it can be expressed by the following equation,

$$\begin{aligned} X_{1(e)} &= X_{2(e)} \\ X_{M-1(e)} &= X_{M(e)} \end{aligned} \tag{3.2.7.15}$$

Discrete Cosine Transform (DCT)

Final step of cepstral analysis is to apply the DCT function to the compressed & weighted filter coefficients. Moreover, it is to de-correlate the filter coefficients. Suppose, the order

of GTCC is 'p', then feature vector will comprise 'p' coefficients. Mathematically, it can be expressed with the following equation,

$$GTCC_k = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} X_{m(\ln+e)} \cos(\frac{\pi k(m-0.5)}{M}) \qquad where,\ 1 \leq k \leq p \quad (3.2.7.16)$$

where, k represents the k^{th} Gammatone Cepstral Coefficient.

3.2.8 Wavelet Analysis

Wavelet gives both time and frequency resolution of a signal. It is a small wave packet, which is duration bound and has concentrated energy in time. Although, these wave packets or wavelets can be shifted in time. Wavelet functions have two properties; scaling and shifting. Scaling is the compression and rarefaction of wavelet i.e. equivalent to variations in frequency of a wave packet whereas translation is the shifting of a wave packet along time axis. These two properties helps in gaining both time and frequency information of a signal. At low frequencies, it gives poor time resolution along with good frequency resolution and vice versa. Two wavelet functions have been used in this study for feature extraction i.e. Daubechies and Symlets. Order of both the wavelet functions were kept to three because characteristics of the source resides in low frequency regions. Moreover, symlet is a modified version of daubechies wavelet and the significant difference between the two functions is that higher-order symlets are symmetric compared to daubechies. Following equation represents general form of the discrete wavelet function.

$$\Psi_{m,n}(t) = \frac{1}{\sqrt{S_0^m}} \Psi \left(\frac{t - nt_0 s_0^m}{s_0^m} \right) \qquad (3.2.8.1)$$

where, m and n are the scaling and time shifting factors, respectively. Following equation is used to calculate detail coefficients of a signal using wavelets.

$$\psi_{m,n}(t) = \int_{-\infty}^{\infty} x(t) \Psi_{m,n}(t) \partial t \qquad (3.2.8.2)$$

3.3 Classification Techniques

For automated detection of objects based on signals acquired via sonar platform, template classifier and variants of neural network have been used in this study. To perform efficient classification, five classifiers have been used, namely, a hyperplane classifier i.e. multilayer layer perceptron with fixed and variable step size (MFNN and VLR-MFNN), a kernel based classifier i.e. radial-basis function neural network (RBF-NN), convolutional neural network (CNN) and a template classifier i.e. dynamic time warping (DTW). Details of each classifier have been discussed in the following.

3.3.1 Neural Networks

Artificial neural networks are inspired with human nervous system i.e. by real biological neurons. First neural network model was developed by McCulloch and Pitts [47]. Since then, thorough usage of neural network classifiers in the field of pattern recognition and signal processing has led to the development of several neural network architectures and learning algorithms. They comprise of associative memory networks i.e. Hopfield memory, Bi-directional associative memory (BAM) and pattern recognition networks i.e. multilayer perceptron (MLP), counter propagation networks etc. Artificial Neural Network is a highly interconnected network of data processing units known as artificial neurons. The connections joining nodes to one another are termed as weights [32]. Moreover, weights are deemed as the single most important factor in determining the output of a neuron. The higher the weight of a link to the neuron, stronger will be the effect of the input applied to that neuron. For a complex network of hundreds of neurons, algorithms are used for weight updating. The process is termed as training. Depending upon the type of weights, the learning and adjustment mechanisms are different. Another important factor is the activation function, which is used to scale the output of the neuron. Artificial neuron acts as a stimulator to the next neuron, as output of one neuron activates another neuron in the next layer. Moreover, all neuron take decisions based on some activation function. A typical neuron with input, output and activation function is illustrated in figure below.

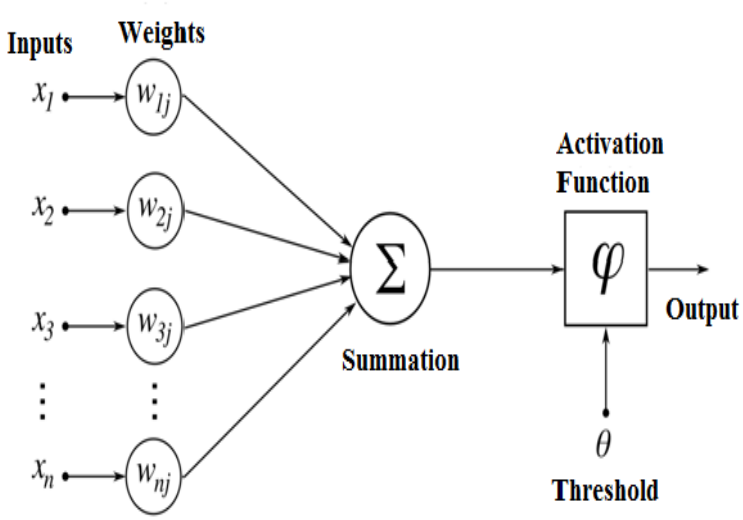

Figure 3.16: A Simple Perceptron

Further, three types of mechanisms exist for training of a classifier, namely, supervised training, un-supervised training and combined supervised/ un-supervised training. Classifiers trained with supervised methods require information of correct class labels of the data set whereas classifiers with un-supervised learning mechanism i.e. vector quantization and clustering, groups un-labelled data into internal clusters. Classification with combined supervised/ un-supervised learning mechanism usually uses clustering to segregate data into internal clusters. Then, the generated clusters are assigned class labels, which is then used for training as done in supervised learning mechanisms.

Why Neural Networks? Artificial Neural Network is a network of massively parallel neurons i.e. distributed processors interconnected to each other in a mesh. It has a tendency of naturally storing information to be used then and forth. ANNs have been applied to a lot of the real world problems of complex nature, making them a lucrative solution compared to current technologies. Neural networks have an ability to derive meanings and finding patterns from an imprecise data that is difficult for an ordinary human being or machine to detect or notice. A trained neural network is an expert system when it comes to analysing the kind of information it is trained with. Neural networks have several advantages, few of them are discussed below,

- **Adaptive:** Ability to learn and adapt on how system should respond to input data based on training or initial experience.

- **Self-Organization:** An ANN has an ability to create its own interpretation or organization of information received during learning time.

- **Real Time Operation:** Computations can be carried out in parallel for an ANN; special type of hardware can be designed and manufactured to utilize this feature of ANN.

- **Fault Tolerance via Redundant Information Coding:** Damage to a neural network could cause performance degradation. However, the network is massively interconnected, so even in case of damage; some of the network capabilities may be retained.

3.3.2 Multilayer Feed-forward Neural Network

Multilayer perceptrons are feed-forward networks which can be trained using back-propagation algorithm [68]. The network comprised of an input layer, an output layer and one or more hidden layers. Number of neurons in the input layer depends upon the size of the feature vector whereas the number of neurons in output layer depends upon the size of the vector representing class labels. Figure 3.17 illustrates a multilayer feed-forward neural network.

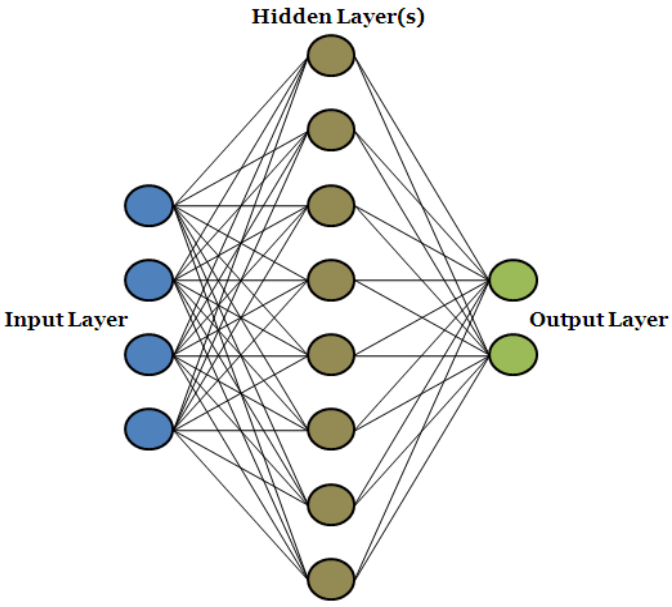

Figure 3.17: Multilayer Feed-forward Neural Network

In back-propagation algorithm, there are two steps, a forward step and a backward step. During the forward step; neuron sends its signal in forward direction. A non-linear activation function is used for decision making at each layer i.e. it transforms an input signal into an output signal. While during the backward step; the error (difference between desired value and output value from the network) is back-propagated to adjust the weights of the links to reduce the accumulated error. Objective of this training phase is to adjust weights in such a way that it reduces the overall error. When error is reduced to its minimum or it reaches an acceptable steady state value the training phase is complete. Back-propagation algorithm is a supervised learning algorithm, therefore, requires every input to be mapped to a class label or output. New set of weights are iteratively calculated based on the amount of error until an overall minimum error is achieved. The mean square error is the objective function and a measure of global error and it is defined as,

$$MSE(n) = \sum_{j=1}^{N_o} (d_j(n) - y_j(n))^2 = \sum_{j=1}^{N_o} e_j^2(n) \tag{3.3.2.1}$$

where, $e_j(n)$ represents the error between the output of the j^{th} neuron and the corresponding desired value. N_0 represents the number of output nodes.

Back-propagation Algorithm - As discussed, Multilayer perceptron uses back-propagation algorithm for training. Its basic steps and weight update mechanism is briefly discussed below,

- The input signal, $x_1, x_2, x_3, \ldots, x_n$, propagates through the input layer neurons to the hidden and output layer neurons in forward direction.
- The error, $e_1, e_2, e_3, \ldots, e_m$, is back propagated from output to the input layer.
- Weight of the i^{th} neuron of the input layer to the j^{th} neuron of the hidden layer is denoted by w_{ij}.
- Weight of the j^{th} neuron of the hidden layer to the k^{th} neuron of the output layer is denoted by w_{jk}
- Error at the k^{th} neuron of the output layer on p^{th} iteration can be expressed as,

$$e_k(p) = d_k(p) - y_k(p)$$

where, $d_k(p)$ and $y_(p)$ represents the desired value and output value at k^{th} neuron on p^{th} iteration, respectively.

- The weight update equation is given by,

$$w_{jk}(p+1) = w_{jk}(p) + \triangle w_{jk}(p)$$

where, $\triangle w_{jk}(p)$ represents the weight adjustment factor.

- Further, weight adjustment factor for the link connecting j^{th} neuron to the k^{th} neuron of the next layer on p^{th} iteration can be mathematically expressed as,

$$\triangle w_{jk}(p) = \alpha y_j(p)\delta_k(p)$$

where, α is the learning rate, $\delta_k(p)$ represents the local gradient at k^{th} neuron of the output layer on p^{th} iteration.

- The local gradient is the product of the gradient of the activation function and the error at the neuron output. The gradient at k^{th} neuron on p^{th} iteration is given by,

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial \mathbf{x}_k(p)} e_k(p)$$

where, $\mathbf{x}_k(p)$ is the weighted input to the k^{th} neuron on p^{th} iteration.

- Moreover, correction of weights, w_{ij}, connecting neurons of input layer to the hidden layer will be made using the same process as above.

3.3.3 Variable Learning Rate Feed-forward Neural Network (VLR-NN)

For variable learning rate - feed-forward neural network, a modified back-propagation algorithm is used for training purpose. For modified back-propagation algorithm, weights associated with i^{th} neuron of the $(l-1)^{th}$ layer to the j^{th} neuron in the l^{th} layer is updated as below:

$$\underbrace{w_{ji}^{(l)}(n+1)}_{\text{New Weights}} = \underbrace{w_{ji}^{(l)}(n)}_{\text{Old Weights}} + \eta(n) \quad \underbrace{\delta_j^{(l)}(n)}_{\text{Local Gradient}} \quad y_i^{(l-1)}(n) \tag{3.3.3.1}$$

where, $\eta(n)$ represents time-varying learning rate. It is updated according to the mechanism of variable step-size learning algorithm in [3] as,

$$\eta(n+1) = \begin{cases} \eta_{max} & if, \rho(n+1) > \eta_{max} \\ \eta_{min} & if, \rho(n+1) < \eta_{min} \\ \rho(n+1) & otherwise \end{cases} \tag{3.3.3.2}$$

Value of $\rho(n+1)$ depends upon the energy of the instantaneous error, that is,

$$\rho(n+1) = \alpha\eta(n) + \gamma e^2(n) \tag{3.3.3.3}$$

Such that, $0 < \alpha < 1$ and $\gamma > 0$. $\delta_j(n)$ in equation 3.3.3.1 represents the local gradient associated with the j^{th} neuron of the l^{th} layer and can be calculated as follows,

$$\delta_j^{(l)}(n) = \begin{cases} (d_j(n) - y_j^{(L)}(n))\phi_j\prime(v_j^{(L)}(n)) & , OutputLayer \\ \phi_j\prime(v_j^{(L)}(n)) \sum\limits_{\forall K} \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n) & , HiddenLayer \end{cases} \tag{3.3.3.4}$$

3.3.4 Radial-Basis Function Neural Network (RBF-NN)

Radial-basis function neural network (RBF-NN), also known as the kernel classifier, has one hidden layer. Kernel function (i.e. usually the Gaussian function) defines the composition of neurons in the hidden layer. Kernel function forms a complex decision space by localizing regions in the space of the transformed data [4] [14]. The only hidden layer uses the non-linear activation function that transforms the data from a low-dimensional space to a high-dimensional space while making it linearly separable. Radial-Basis function is a term which depicts the use of a kernel function as a non-linear activation function in the hidden layer. Each hidden layer neuron compares each of the link inputs coded in form of weights. The activation function generates a strong output when input is near to the centroid of the kernel function in terms of Euclidean distance and as the input moves away from the centroid, output signal decreases monotonically. Further, Euclidean distance is the metric used for distance calculation. Radial-basis function neural network (RBF-NN) classifier is a feed-forward mapping network that has a lot of applications in classification problems related to object identification based on speed and underwater transients. It provides an alternative tool to learn in neural networks. The advantage of RBF is that it does not get trapped into local minima or maxima of the objective function surface whereas the disadvantage is that it requires a large space for input representation as number of centers depends upon the distribution

of input data. Here, the main objective is to design an artificial neural network with good adaptation and inference ability along with having fewer number of processing units to avoid long taxing calculations. In this work, centroids of the RBF network were determined from clustering while width was kept constant during training. Figure 3.18 illustrates a typical RBF network.

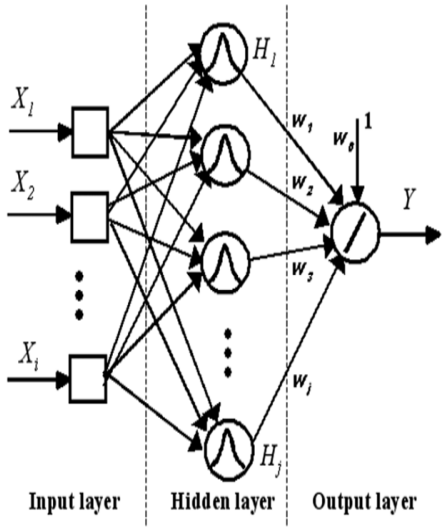

Figure 3.18: Radial-Basis Function Neural Network

The main factors of RBF are,

1. RBF-Networks are used for performing complex (non-linear) pattern classification and function estimation tasks.

2. They are called single layer feed-forward neural networks.

3. Nodes on hidden layer uses an activation function.

4. Output nodes only sum up the weighted input and doesn't use any activation function in decision making.

5. Weights are updated from input to hidden and then, from hidden to output layer.

6. It takes probably more computation time compared to MLP but learning is faster compared to MLP.

7. Larger number of hidden layer neurons are required compared to MLP.

The interpolation of $N-$points in a $D-$dimensional space requires input vectors, X_p, of $D-$length as

$$\mathbf{x^p} = \{x_i^p; \qquad i = 1, 2, 3, \ldots, D\} \tag{3.3.4.1}$$

The input vectors are mapped to respective target outputs, t^P, goal is to calculate function, $f(x)$, such that,

$$f(\mathbf{x}^p) = t^p \qquad \forall p = 1, 2, \ldots\ldots, N \tag{3.3.4.2}$$

RBF approach introduces N basis functions for each $D-$dimensional data point in the input set. Each data point will be mapped to one of the basis which is closest to it in terms of distance. The function is of the form $\phi(||\mathbf{x} - \mathbf{x}^P||)$, where, $\phi(.)$ is some non-linear function i.e. Gaussian function etc. Moreover, the output can be mathematically written as,

$$f(\mathbf{x}) = \sum_{p=1}^{N} w_p \phi(||\mathbf{x} - \mathbf{x^p}||) \tag{3.3.4.3}$$

Solving equation 3.3.4.2 and 3.3.4.3 will yield appropriate value of weights with respect to the transformed input set. It is given by,

$$f(\mathbf{x^q}) = \sum_{p=1}^{N} w_p \phi(||\mathbf{x^q} - \mathbf{x^p}||) = t^q \tag{3.3.4.4}$$

Further, the matrix form representation is, let $f(\mathbf{x}^q) = t^q$, $\mathbf{w} = \{w_p\}$ and $\Phi = \{\phi_{pq} = \phi||\mathbf{x}^q - \mathbf{x}^p||\}$. Then, equation 3.3.4.4 becomes,

$$\Phi^T w = t \tag{3.3.4.5}$$

3.3.5 Dynamic Time Warping (DTW)

Dynamic time warping (DTW) [59] is a template classifier. It is a measure for calculating similarity between time varying sequences. It has been used long and to greater extent in applications of speech signal processing. Its distinguishing feature is that it performs non-linear mapping by minimizing distance between two sequences. Non-linear (elastic) mapping is a good similarity measure as sequences out of phase but similar in shape or pattern are easier to match. Usually, Euclidean distance is the metric used for distance calculation. Following figure illustrates non-linear mapping of one signal on another.

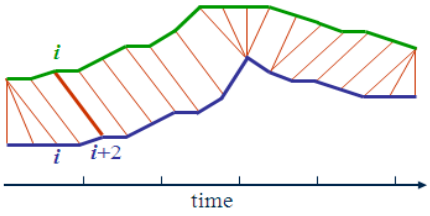

Figure 3.19: Non-linear Mapping of One Signal on Another [69]

To find the best alignment between signal A and B, one needs to find the path through the grid, $P = p_1, p_2, \ldots, p_s, \ldots, p_k$, where, $p_s = (i_s, j_s)$. Which minimizes the total distance between two signals. Here, P is called a warping function. Following figure illustrates the warping function,

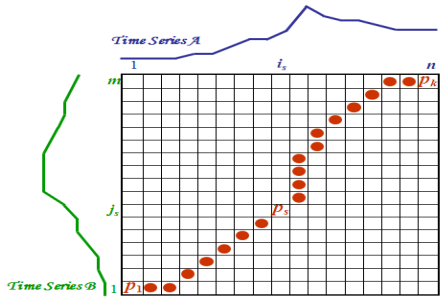

Figure 3.20: Depiction of Warping Function [69]

Time-normalized distance between A and B is given by,

$$D(A,B) = \left[\frac{\sum_{s=1}^{k} d(p_s).w_s}{\sum_{s=1}^{k} w_s} \right] \tag{3.3.5.1}$$

where, $d(p_s)$ represents distance between points i_s and j_s and w_s represents the weighting coefficient. Moreover, the best alignment path between A and B is given by,

$$P_o = \arg_p \min(D(A,B)) \tag{3.3.5.2}$$

However, the number of possible warping paths through the grid are exponentially explosive, so optimization measures available in literature [69] are taken to reduce the computational complexity of the process. They are listed below along with the figure 3.21 illustrating effects these will have on the computational complexity.

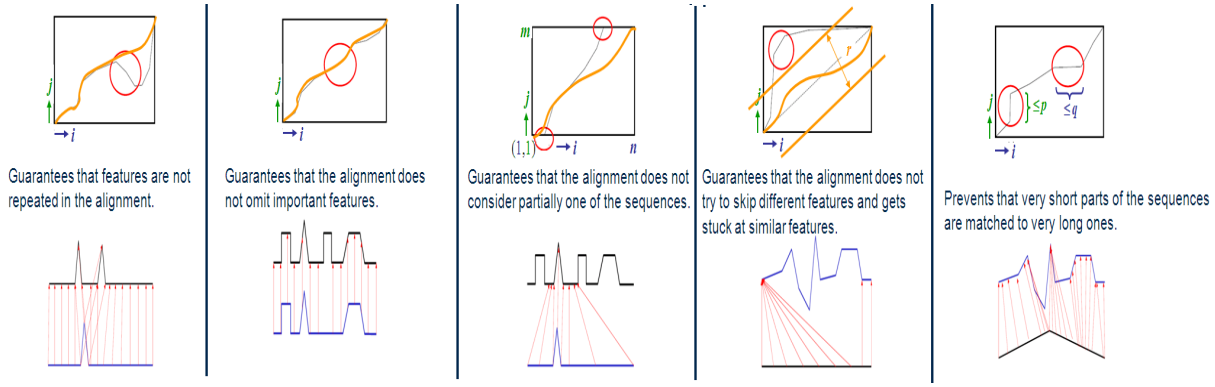

Figure 3.21: Optimization Methods **(1)** Monotonicity **(2)** Continuity **(3)** Boundary Conditions **(4)** Warping Window **(5)** Slope Constraint [69]

1. Monotonicity: The alignment path does not go back in "time" index, that is, $i_{s-1} \leq i_s$ and $j_{s-1} \leq j_s$.

2. Continuity: The alignment path does not jump in "time" index, that is, $i_s - i_{s-1} \leq 1$ and $j_s - j_{s-1} \leq 1$.

3. Boundary Conditions: The alignment path starts at the bottom left and ends up at the top right, that is, $i_1 = 1, i_k = n$ and $j_1 = 1, j_k = m$.

4. Warping Window: A good alignment path is unlikely to wander too far from the diagonal, that is, $|i_s - j_s| \leq r$, where $r > 0$ is the window length.

5. Slope Constraint: $(j_{sp} - j_{so})/(i_{sp} - i_{so}) \leq p$ and $(i_{sq} - i_{so})/(j_{sq} - j_s) \leq q$, where $q \geq 0$ are the number of steps in $x-$direction and $p \geq 0$ are the number of steps in $y-$direction. After q steps in x direction one must step in y direction and vice versa, that is, $S = p/q \in [0, \infty]$.

Algorithm - Steps of Dynamic Time Warping algorithm are discussed below,

1. **Initial condition:** $g(1, 1) = d(1, 1)$.

2. **DP-equation:**

$$g(i,j) = \min \begin{cases} g(i,j-1) + d(i,j) \\ g(i-1,j-1) + d(i,j) \\ g(i-1,j) + d(i,j) \end{cases} \tag{3.3.5.3}$$

3. **Warping window:** $j - r \leq i \leq j + r$

4. **Time-normalized distance:**

$$D(A,B) = g(n,m)/C \tag{3.3.5.4}$$

where, $C = n + m$

3.4 Dimensionality Reduction Techniques

Computational complexity of a system is proportional to the size of the input vector. Features extracted using Low Frequency Analysis & Ranging (LOFAR) method have been used to study the effects of dimension reduction on classification results. For the purpose of dimensionality reduction, principal component analysis (PCA) and linear discriminant analysis (LDA) have been used. Details of both the methods are in the following.

3.4.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) [39] is a statistical data analysis technique. It is also termed as an un-supervised classifier and known for increasing the variance amongst the data elements in the input set. It has greater applications in fields of pattern recognition and cryptography. It is very commonly used in applications where dimension of the data set is to be reduced. It transforms the data set in high-dimensional space into a low-dimensional space reducing the overall processing cost with out losing much information.

It is used for finding similarities and patterns present in data. Since, there is less chance of efficiently analysing data of higher dimensions analytically, that's where PCA

helps the most; it is a powerful tool that helps in analysing data of higher dimension i.e. finding similarities and differences. One of the most important features of PCA that makes it ever so useful is that once patterns are identified in data, it can be compressed.

Steps for implementation of Principal Component Analysis

1. First step of principal component analysis (PCA) is to obtain zero-mean data, that is, because we are only interested in patterns present in the data not the absolute values. Subtracting mean value from data will yield zero-mean data.

2. Covariance matrix is calculated using the zero-mean input data. Result is a square matrix, whose dimension is same as that of the input data.

3. Next step is the calculation of Eigenvectors and Eigenvalues using covariance matrix. Eigenvectors are arranged based on their respective Eigenvalues. Eigenvector with the highest Eigenvalue will be placed first and will be referred to as the principal component. Further, Eigenvectors are normalized to unity as absolute values don't have any information only patterns do.

4. Final step is to transform data using Eigenvectors. Eigenvectors acts as basis of the space in which the transformed data lies. Usage of all components in data transformation yields data with maximum variance or spread. Leaving out Eigenvectors from the set of principal components during data transformation will result in loss of information and it will also weaken the discriminating ability of the feature set.

3.4.2 Linear Discriminant Analysis (LDA)

It is a statistical data analysis technique, also known as a supervised classifier. It transforms data elements into an Eigen space while maximizing the projected variance between data elements belonging to different classes and reducing the intra-class projected variance i.e. minimizing the variance between data elements belonging to same class [70]. Further, key difference between PCA and LDA is that the lateral deals with labelled

data. Objective function for LDA can be mathematically expressed as,

$$\max_{\mathbf{w}} J(w) = \frac{\mathbf{w}^T B_w}{\mathbf{w}^T B_s} \tag{3.4.2.1}$$

where, B_w and B_s represents the between and within class scatter matrices, respectively. Main goal of LDA is to find a vector $\mathbf{w}$ that maximizes the objective function.

3.5 Deep Learning and Multi-linear Subspace Learning

Deep learning and sub-space learning have shown to outperform many standard and conventional methods in acoustic modelling in many research applications related to speech and acoustics [37] [21] but there haven't been many studies available investigating their performance for under water acoustics. So, apart from all the front-ends and back-ends discussed in previous sections, two relatively newer approaches have also been used in this study. Amongst the two approaches, one has been used for feature learning and the other being used for classification. For classification, convolutional neural network (CNN), a deep learning approach has been used whereas multi-linear principal component analysis (MPCA), a multi-linear sub-space learning approach has been used for feature learning and dimensionality reduction. CNN and MPCA have both been applied successfully to problems where conventional techniques have failed to perform. Moreover, both techniques have worked well in environments that uses tensor objects as inputs i.e 2D or 3D Matrices. Further, both techniques have given good and sustained performance levels where size, dimension and complexity of data is large. Brief details of both the techniques are in the following.

3.5.1 Convolutional Neural Network (CNN)

It is a deep learning approach, which has worked well in classification problems where data is tensor, complex and separability is difficult. convolutional neural networks (CNNs) are inspired with visual mechanism in living beings. The visual cortex in human brain contains a lot of cells that are sensitive to light and detects light in small overlapping sub-regions in the visual fields known as the receptive fields. Complex cells have larger visual fields. Moreover, these cells act as filters in realizing the convolution

operation. Convolutional neural network comprised of two types of layers, namely, the convolutional layer and the sub-sampling or pooling layer. One major advantage of the convolution neural network is the use of shared filter bank (weights) in the convolutional layers. This improves performance while reducing the size of the parameter set [1]. After the aforesaid layers comes the optional fully connected layer(s) that connects output layer with the complete network. Input to the convolution neural network is always a 2D or 3D data matrix. The convolutional layer comprises of 'k' number of filters which convolves with the input matrix to produce feature maps. Further, each map is passed on to the sub-sampling layer for max or mean pooling operations. Either before or after pooling operation, each feature map is passed through a non-linear activation function i.e. sigmoid or ReLU etc. to incorporate non-linearity into the feature maps and as mentioned earlier, an output layer follows the convolutional and sub-sampling layers. Moreover, back-propagation algorithm has been used for weights correction, which has already been discussed in previous section.

The convolution operation in the convolutional layer can be mathematically described by the following equation,

$$CFM_j^i = f(\sum_{i=1}^{I} X_i^l \otimes K_{ij}^l + b_j^l) \tag{3.5.1.1}$$

where, 'CFM_j^l' is the output feature map from the l^{th} convolutional layer, 'f' represents the non-linear activation function, 'X_l^l' represents the l^{th} input matrix, 'K_{ij}^l' represents the kernal and 'b_j^l' represents the bias value. Similarly, the pooling operation in the sub-sampling layer can be mathematically described by the following equation,

$$PFM_{i,m}^l = \max_{n-1}^{G} CFM_{i,(m-1)*s+n}^l \tag{3.5.1.2}$$

where, '$PFM_{i,m}^l$' is the output feature map from the l^{th} sub-sampling layer, 'G' represents the pooling size and 's' represents the shifting parameter.

3.5.2 Multi-linear Principal Component Analysis (MPCA)

It is a multi-linear subspace (MSL) learning approach, which is used for feature learning and dimensionality reduction of tensor objects i.e. multi-dimensional objects. It is designed to work with tensors of any order i.e. 1D, 2D, 3D etc. Moreover, the objective

of MPCA is to find Eigen tensors that tends to capture maximum variations present in the input data. Traditionally, linear subspace learning (LSL) mechanisms i.e PCA and LDA, were used a lot and they have mostly been used to reduce the dimensions of the data set. Moreover, almost all LSL methods represent input as a vector and solve for transforming the input vectors into an optimal lower-dimensional space. However, with multi-dimensional data LSL methods haven't been effective because it tends to break the natural structure of the objects which results in loss of information and therefore the transformed data usually tends to lose variations present in the original tensors. Moreover, the author in [45] has discussed all mathematical and algorithmic details of multi-linear principal component analysis (MPCA) related to it's implementation. Also, results presented in [45] signifies the importance and need of usage of the MSL approaches for feature learning and dimensionality reduction of tensor objects.

4 Simulation Results

4.1 Experimental Conditions

In this chapter performance evaluation of all detection and classification schemes have been made using two sets of data, that is, a raw dataset having acoustic samples belonging to 4 different classes of ships and a synthetic dataset, URL: http://www.dosits.org [79], having acoustic samples belonging to 20 different classes of underwater objects i.e. ships and sea species. All scripts and simulations have been written and conducted in MATLAB, respectively. Moreover, toolboxes for convolution neural network (CNN) [53] and multi-linear principal component analysis (MPCA) [45] have been used. Further, Classification results were observed under the effects of noise i.e. additive white Gaussian noise (AWGN). Following are the 20 class labels of acoustic samples downloaded from DOSITS [79],

1. Ship with a Vessel in a Tow
2. Merchant Vessel
3. Commercial Ship
4. SONAR
5. Mantis Shrimp
6. Mantis Shrimp - Patek Caldwell - In presence of an Intruder
7. Torpedo

8. Blue Grunt (Fish)
9. Toad-fish
10. Silver Perch
11. Blue Whale
12. Blue Whale - Sulphur Bottom Whale
13. Amazon River Dolphin
14. Gray Whale
15. Sperm Whale (Coda)
16. Sperm Whale (Creak)
17. Atlantic Croaker
18. Pilot Whale
19. Harbour Porpoise
20. Surveillance Towed Array Sensor System - Low Frequency Active (SURTASS-LFA) SONAR

After feature extraction process, acquired feature set was fed to the classifier for automated recognition. The task of the recognition system is to correctly identify the class label of the sample under test. Moreover, half of the generated samples were used for training while the remaining half was used to test the performance of the trained network. All samples were chopped into ensembles of about 25 ms of duration to mitigate the effects of non-stationarity. Moreover, consecutive frames had an overlapping of around 50% in all feature extraction front ends except for Bartlett spectral estimation method where overlapping between frames is not required. Feature vectors were calculated for each frame of the acoustic samples while using all the detection schemes as

discussed in chapter 3. Moreover, each feature vector was represented using ***'p'*** number of coefficients.

A profile for % recognition is created for each front-end feature extraction method with every back-end classification scheme. Simulations were repeated at different values of signal to noise ratio. Five renowned classifier were used for classification of objects based on extracted feature set. Namely, multilayer feed-forward neural network (MFNN), variable learning rate feed-forward neural network (VLR-NN), radial-basis function - neural network (RBF-NN), convolutional neural network (CNN) and dynamic time warping (DTW). Effects of dimensionality reduction on classification results were also observed. Three techniques were used for the said purpose i.e. two linear subspace learning schemes, namely, principal component analysis (PCA) and linear discriminant analysis (LDA) and a multi-linear subspace learning (MSL) approach, multi-linear principal component analysis (MPCA). Details of simulation environment and parameters are listed in table 4.1 below.

Table 4.1: Simulation Details

Parameters	Attributes
Sampling Frequency (Hz)	44100
Down-Sampling Factor	5
Sample Size (FrameSize) (sec)	$25ms$
Step-Size (sec)	$12.5ms$
Database	DOSITS [79]
	Raw Dataset
No. of Samples in each Class - DOSITS [79]	21200 Approx
No. of Samples in each Class - Raw Dataset	657600 Approx
No. of Classes	DOSITS - 20
	Raw Dataset - 4
Percentage of Samples for Training	50%
Percentage of Samples for Testing	50%
Noise Type	AWGN
SNR Range	$-20dB, -10dB, 0dB, 10dB$ and $10dB$
Front-End Models	LOFAR, DEMON, LPC, LPCC,
	PLP, MFCC, GFCC, Wavelet and MPCA
Back-End Units	MFNN, VLR-MFNN, RBF-NN, DTW and CNN
Dimension Reduction Techniques	PCA and LDA
Platform	MATLAB

4.1.1 Graphical User Interface Model

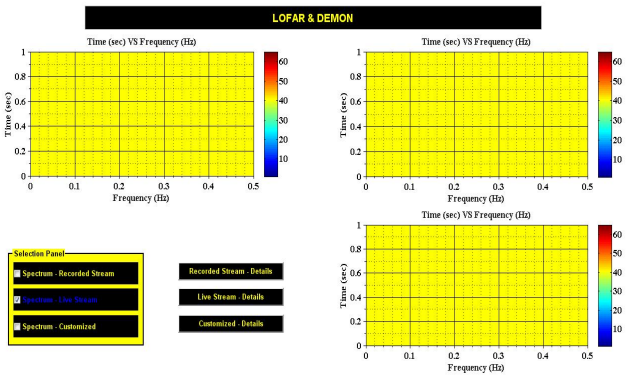

Figure 4.1: Graphical User Interface Model for Detection of Envelope Modulation on Noise (DEMON) and Low Frequency Analysis and Ranging (LOFAR) Analysis

In addition, a graphical user interface (GUI) has been developed for performing two analysis, that is, detection of envelope modulation on noise (DEMON) and low frequency analysis & ranging (LOFAR), on the acquired acoustic sample. GUI was build using MATLAB. Figure 4.1 illustrates the graphical user interface for the aforementioned system. The model can be used to perform LOFAR and DEMON analysis on recorded and real-time streams. Figure 4.2 gives view of the GUI model illustrating signal spectra after DEMON/ LOFAR analysis on the real-time stream. For real-time streams, data acquisition was made after every 25 ms.

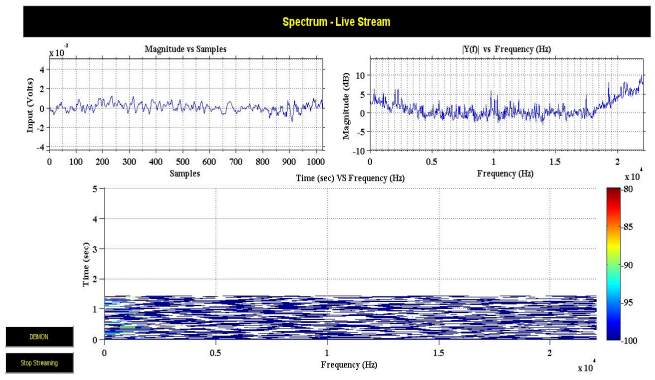

Figure 4.2: Graphical User Interface Model for Signal Analysis - Real-time Stream

Figure 4.4 and 4.3 illustrates the graphical user interface for performing LOFAR & DEMON signal analysis on the recorded sequences and for generating synthetic spectra based on source characteristics, respectively.

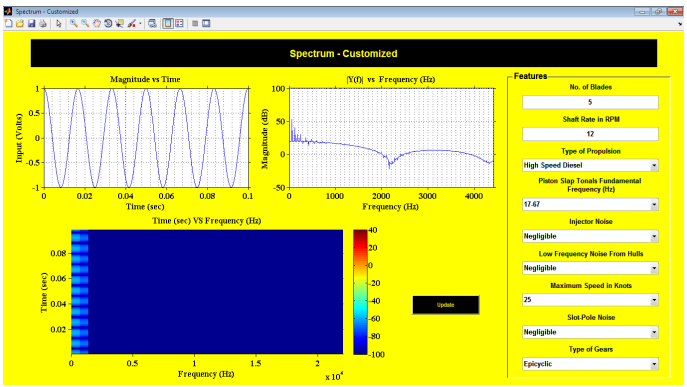

Figure 4.3: Graphical User Interface Model for Generating Synthetic Spectra

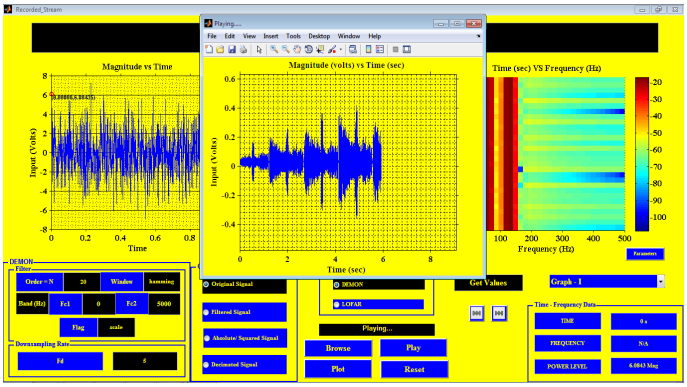

Figure 4.4: Graphical User Interface Model for Signal Analysis - Recorded Stream

4.2 Detection and Classification

As discussed earlier in section 4.1, different signal detection and classification techniques have been used in this study. First, feature extraction is performed and then obtained feature set is fed to the classifier for classification. Moreover, feature sets acquired using aforementioned approaches were first normalized before being fed to the classifier as inputs. Table 4.2 give details of the parameters for the respective classification techniques.

Table 4.2: Classification Techniques - Parameters

Technqiues	Parameters	Values
MFNN	**Activation Functions,** ϕ	Tansig and Logsig
	No. of Layers	05
	No. of Hidden Layers	03
	No. of Neurons in Input Layer	As the size of the FV
	No. of Neurons in Hidden Layer	150-100-100
	No. of Neurons in Output Layer - DOSITS	20
	No. of Neurons in Output Layer - Raw Dataset	4
	Learning Rate, η	0.05
	Total Epochs	500
VLR-MFNN	**Activation Functions,** ϕ	Tansig and Logsig
	No. of Layers	05
	No. of Hidden Layers	03
	No. of Neurons in Input Layer	As the size of the FV
	No. of Neurons in Hidden Layer	150-100-100
	No. of Neurons in Output Layer - DOSITS	20
	No. of Neurons in Output Layer - Raw Dataset	4
	Learning Rate, η	0.008 to 0.05 (Variable)
	Alpha, α	0.2
	Gamma, γ	0.00001
	Total Epochs	500
RBF-NN	**RBF Type**	Recursive RBF
	No. of Layers	03
	No. of Hidden Layers	01
	No. of Neurons in Input Layer	As the size of the FV
	No. of Neurons in Hidden Layer	20
	No. of Neurons in Output Layer - DOSITS	20
	No. of Neurons in Output Layer - Raw Dataset	4
	Learning Rate, η	0.09
	Total Epochs	100
	Non-Linear Function, ϕ	Gaussian Function
	Centres	Fixed Centres - Using Euclid. Dis.
	Variance, σ	Fixed Variance - 0.2

Table 4.3 and Table 4.4 give details of the simulation results for both the datasets. Every detection technique has been used with every classification technique except for CNN and MPCA. Moreover, all presented results have been averaged over at least 10 iterations. Further, all simulations have been performed under the influence of noise at different levels of signal-to-noise ratio (SNR).

Table 4.3: DOSITS - %Recognition - Feature Extraction Techniques vs Classification Techniques

			Classification Techniques			
		SNR	MFNN	VLR-NN	RBF-NN	DTW
Feature Extraction Techniques	**DEMON**	−20	-	-	-	81.75%
		−10	-	-	-	84.58%
		0	69.00%	68.02%	71.05%	91.21%
		10	90.38%	89.25%	97.56%	94.53%
		20	98.42%	97.35%	99.84%	99.57%
	DEMON (LPF)	−20	-	-	-	89.00%
		−10	-	-	-	90.21%
		0	51.21%	50.68%	59.04%	93.24%
		10	79.87%	78.38%	98.82%	98.34%
		20	93.90%	92.82%	94.53%	99.53%
	DEMON (HILBERT)	−20	-	-	-	86.25%
		−10	-	-	-	88.34%
		0	58.49%	59.85%	58.83%	91.23%
		10	81.50%	81.08%	93.30%	94.87%
		20	95.06%	93.06%	98.27%	99.12%
	LOFAR	−20	-	-	-	86.21%
		−10	-	-	-	88.98%
		0	75.38%	74.41%	73.17%	93.57%
		10	91.38%	89.33%	95.19%	95.33%
		20	98.71%	97.21%	97.66%	99.10%
	LOFAR (Welch)	−20	-	-	-	22.25%
		−10	-	-	-	53.00%
		0	56.60%	55.60%	57.20%	95.75%

continued on next page...

... Table 4.3 continued

Feature Extraction Techniques		Classification Techniques				
		SNR	MFNN	VLR-NN	RBF-NN	DTW
		10	76.00%	73.00%	75.30%	96.32%
		20	89.40%	87.20%	87.20%	98.23%
	LOFAR (Bartlett)	−20	-	-	-	23.25%
		−10	-	-	-	51.00%
		0	56.20%	54.20%	54.80%	89.75%
		10	77.49%	74.40%	76.80%	94.23%
		20	87.80%	92.40%	91.10%	98.21%
	LPA (Bareiss Method)	−20	-	-	-	80.00%
		−10	-	-	-	89.00%
		0	28.43%	28.95%	22.07%	92.23%
		10	51.87%	55.10%	54.18%	95.21%
		20	73.46%	75.54%	76.33%	99.43%
	LPA (Durbin's Method)	−20	-	-	-	89.75%
		−10	-	-	-	91.22%
		0	26.41%	26.41%	21.23%	93.45%
		10	45.59%	48.61%	50.32%	96.32%
		20	74.49%	76.56%	76.62%	98.45%
	LPCC	−20	-	-	-	92.00%
		−10	-	-	-	93.23%
		0	76.64%	76.06%	69.63%	95.54%
		10	95.14%	75.30%	92.62%	98.21%
		20	99.53%	97.56%	99.74%	99.43%
	PLP-BFCC	−20	-	-	-	16.25%
		−10	-	-	-	48.75%
		0	57.38%	56.86%	55.31%	89.50%
		10	89.31%	88.23%	87.26%	95.21%
		20	98.77%	98.13%	98.79%	99.34%
	MFCC	−20	-	-	-	85.00%
		−10	-	-	-	89.23%

continued on next page...

... Table 4.3 continued

Feature Extraction Techniques			Classification Techniques			
		SNR	MFNN	VLR-NN	RBF-NN	DTW
		0	67.80%	67.93%	53.89%	94.21%
		10	91.31%	91.73%	92.88%	97.34%
		20	98.24%	98.13%	03.31%	99.11%
	GTCC	−20	-	-	-	09.25%
		−10	-	-	-	25.00%
		0	62.27%	61.72%	61.51%	78.00%
		10	91.20%	91.64%	89.62%	89.32%
		20	98.92%	98.75%	99.55%	97.67%
	Wavelet (Daubechies)	−20	-	-	-	89.23%
		−10	-	-	-	91.65%
		0	27.69%	21.18%	42.14%	92.43%
		10	74.49%	65.95%	95.48%	95.12%
		20	74.48%	93.27%	99.16%	99.32%
	Wavelet (Symlets)	−20	-	-	-	82.21%
		−10	-	-	-	87.44%
		0	25.35%	23.15%	35.89%	91.24%
		10	54.07%	59.12%	93.25%	96.21%
		20	52.65%	88.41%	99.50%	99.46%

Table 4.4: Raw Dataset - %Recognition - Feature Extraction Techniques vs Classification Techniques

			Classification Techniques			
		SNR	MFNN	VLR-NN	RBF-NN	DTW
	DEMON	−20	-	-	-	68.57%
		−10	-	-	-	89.87%
		0	41.48%	41.34%	62.93%	92.31%
		10	63.18%	71.20%	99.12%	97.23%

continued on next page...

... Table 4.4 continued

Feature Extraction Techniques		Classification Techniques				
		SNR	MFNN	VLR-NN	RBF-NN	DTW
	DEMON (LPF)	20	88.12%	95.80%	99.46%	99.32%
		−20	-	-	-	24.64%
		−10	-	-	-	76.96%
		0	37.30%	38.29%	43.98%	86.27%
		10	51.07%	46.71%	97.75%	95.24%
		20	64.21%	53.50%	99.05%	98.21%
	DEMON (HILBERT)	−20	-	-	-	33.57%
		−10	-	-	-	89.11%
		0	42.02%	42.30%	54.41%	93.12%
		10	62.88%	63.05%	99.93%	96.12%
		20	79.86%	93.75%	100%	98.45%
	LOFAR	−20	-	-	-	85.32%
		−10	-	-	-	89.76%
		0	42.77%	41.25%	56.66%	91.43%
		10	64.96%	67.63%	99.95%	98.23%
		20	82.68%	94.45%	100%	99.24%
	LOFAR (Welch)	−20	-	-	-	04.43%
		−10	-	-	-	04.57%
		0	93.71%	88.86%	91.64%	12.86%
		10	96.14%	89.29%	99.64%	39.71%
		20	96.43%	92.86%	100%	90.86%
	LOFAR (Bartlett)	−20	-	-	-	04.43%
		−10	-	-	-	06.00%
		0	91.29%	85.86%	92.29%	11.43%
		10	85.14%	89.00%	99.86%	42.43%
		20	89.29%	85.71%	100%	96.29%
	LPA (Bareiss Method)	−20	-	-	-	55.36%
		−10	-	-	-	89.56%
		0	38.50%	37.30%	35.96%	92.37%

continued on next page...

... Table 4.4 continued

Feature Extraction Techniques		Classification Techniques				
		SNR	MFNN	VLR-NN	RBF-NN	DTW
		10	51.66%	47.38%	59.96%	96.34%
		20	70.61%	67.43%	95.68%	99.45%
	LPA (Durbin's Method)	−20	-	-	-	88.21%
		−10	-	-	-	91.34%
		0	39.79%	36.77%	35.29%	95.12%
		10	49.63%	52.50%	74.29%	98.56%
		20	68.52%	80.07%	99.32%	99.43%
	LPCC	−20	-	-	-	59.64%
		−10	-	-	-	82.23%
		0	58.39%	55.59%	71.02%	89.68%
		10	81.00%	83.52%	99.88%	94.32%
		20	88.57%	98.14%	100%	99.65%
	PLP-BFCC	−20	-	-	-	52.14%
		−10	-	-	-	85.88%
		0	39.79%	38.95%	53.14%	91.23%
		10	55.02%	51.98%	92.50%	95.12%
		20	72.54%	85.71%	99.25%	99.21%
	MFCC	−20	-	-	-	59.64%
		−10	-	-	-	87.32%
		0	46.38%	46.39%	45.98%	91.34%
		10	67.43%	75.59%	98.86%	95.12%
		20	83.88%	95.11%	100%	98.43%
	GTCC	−20	-	-	-	03.93%
		−10	-	-	-	04.92%
		0	43.66%	43.96%	45.54%	08.75%
		10	63.71%	65.41%	91.07%	34.46%
		20	79.45%	89.84%	99.98%	98.57%
		−20	-	-	-	89.23%
		−10	-	-	-	91.24%

continued on next page...

... Table 4.4 continued

		Classification Techniques			
	SNR	MFNN	VLR-NN	RBF-NN	DTW
Wavelet (Daubechies)	0	30.21%	31.39%	42.20%	94.35%
	10	30.23%	38.07%	66.75%	96.46%
	20	46.25%	67.61%	91.75%	99.23%
Wavelet (Symlets)	−20	-	-	-	87.32%
	−10	-	-	-	91.23%
	0	31.68%	30.34%	31.95%	94.13%
	10	36.63%	35.02%	77.86%	95.87%
	20	40.89%	42.39%	99.95%	98.23%

4.3 Results - Linear Subspace Learning: Dimensionality Reduction vs Classification

Efficiency and feasibility of realization of any system is govern by a lot of parameters i.e. size on chip, number of mathematical operations, processing time, delays, memory requirements and power consumption etc. Moreover, a lot of hardware and software constraints limits the performance of a system. In this study, an objective of building a low-cost processing unit was also undertaken for detection and classification of objects based on acoustic signals. For the said purpose, two linear subspace learning (LSL) techniques, namely, principal component analysis (PCA) and linear discriminant analysis (LDA) have been used for dimensionality reduction of the feature sets and thus, reducing the overall computational cost. Moreover, effects of dimensionality reduction were also observed on classification rates. Feature set acquired via low frequency analysis & ranging (LOFAR) with both Bartlett and Welch method have been used to study the effects of dimensionality reduction on percentage classification. For classification, variable learning rate neural network (VLR-NN) was used having parameters as mentioned in table 4.2. Both the aforementioned datasets have been used to evaluate the performance of the said dimensionality reduction techniques. Feature set acquired for DOSITS [79] and raw acoustic samples had 127 and 63 dimensions, respectively. Effects of dimensionality reduction on classification accuracies were observed by reducing the di-

mensions to 25%, 50% and 75% of original feature set at different levels of signal-to-noise ratio (SNR).

Table 4.5 to 4.8 list classification results for dataset downloaded from DOSITS [79]. Further, dimensions of the feature set were varied and results were observed at different values of signal-to-noise ratio (SNR) i.e. $0dB, 10dB$ and $20dB$.

Table 4.5: DOSITS - Dimension vs % Recognition - Principal Component Analysis (Lofar - Welch)

		Dimensions - Feature Vector			
		64	80	96	127
SNR	0	57.80%	57.60%	58.30%	57.50%
	10	71.60%	72.20%	74.50%	85.50%
	20	85.50%	86.70%	85.50%	82.00%

Table 4.6: DOSITS - Dimension vs % Recognition - Linear Discriminant Analysis (Lofar - Welch)

		Dimensions - Feature Vector			
		64	80	96	127
SNR	0	35.20%	35.30%	36.25%	37.55%
	10	67.45%	68.10%	68.55%	70.00%
	20	82.50%	84.95%	85.50%	87.00%

Table 4.7: DOSITS - Dimension vs % Recognition - Principal Component Analysis (Lofar - Bartlett)

		Dimensions - Feature Vector			
		64	80	96	127
SNR	0	53.00%	55.00%	51.40%	57.80%
	10	72.60%	72.60%	72.00%	75.60%
	20	79.20%	87.40%	78.80%	89.40%

Table 4.8: DOSITS - Dimension vs % Recognition - Linear Discriminant Analysis (Lofar - Bartlett)

		Dimensions - Feature Vector			
		64	80	96	127
SNR	0	55.40%	54.45%	56.80%	59.00%
	10	75.70%	78.00%	79.80%	80.20%
	20	89.70%	91.70%	91.25%	91.65%

Table 4.9 to 4.12 list classification results for dataset acquired via sonar platform. Further, dimensions of the feature set were varied and results were observed at different values of signal-to-noise ratio (SNR) i.e. $0dB, 10dB$ and $20dB$.

Table 4.9: Raw Dataset - Dimension vs % Recognition - Principal Component Analysis (Lofar - Welch)

		Dimensions - Feature Vector			
		16	32	48	63
SNR	0	76.71%	87.57%	82.86%	86.14%
	10	92.57%	88.14%	94.57%	84.57%
	20	89.14%	89.29%	92.86%	96.43%

Table 4.10: Raw Dataset - Dimension vs % Recognition - Linear Discriminant Analysis (Lofar - Welch)

		Dimensions - Feature Vector			
		16	32	48	63
SNR	0	50.75%	63.11%	70.82%	69.71%
	10	89.82%	89.71%	91.82%	90.32%
	20	93.54%	91.14%	91.57%	92.04%

Table 4.11: Raw Dataset - Dimension vs % Recognition - Principal Component Analysis (Lofar - Bartlett)

		Dimensions - Feature Vector			
		16	32	48	63
SNR	0	81.43%	84.43%	83.86%	86.86%
	10	94.14%	84.43%	91.29%	95.00%
	20	92.71%	92.71%	89.14%	85.71%

Table 4.12: Raw Dataset - Dimension vs % Recognition - Linear Discriminant Analysis (Lofar - Bartlett)

		Dimensions - Feature Vector			
		16	32	48	63
SNR	0	58.21%	67.43%	71.29%	75.75%
	10	84.11%	89.89%	92.93%	94.32%
	20	86.79%	91.07%	98.61%	88.32%

4.4 Results - Deep Learning and Multi-linear Subspace Learning

In this study, convolutional neural network (CNN), a deep learning approach has been used for classification and multi-linear principal component analysis (MPCA), a multi-linear subspace learning technique, has been used for dimensionality reduction & feature extraction. In both the techniques, spectrograms (Frequency vs Time or 2D Matrix) have been used as inputs. Moreover, A spectrogram was created for an acoustic window of duration 20 ms with 10 ms of overlapping between two consecutive windows. Moreover, both the time and frequency axes were divided into 98 slices, representing the spectrogram as a 98×98 matrix. Details of parameters for convolution neural network (CNN) and multi-linear principal component analysis are in table 4.13.

Table 4.13: Details of Parameters for CNN and MPCA

Technqiues	Parameters	Values
CNN	**Activation Function,** ϕ	Sigmoid
	Size of the Input (Feature Vector)	98×98
	No. of Hidden Layers	05
	No. of Convolution Layers	02
	No. of Output Maps in Convolution Layers	$20 - 10$
	Size of Kernel	7
	No. of Sub-sampling Layers	02
	Scaling Factor in Sub-sampling Layer	02
	No. of Neurons in Output Layer - DOSITS	20
	No. of Neurons in Output Layer - Raw Dataset	4
	Learning Rate, η	0.01
	Batch Size	7
	Total Epochs	5000
MPCA	**Size of the Input (Feature Vector)**	98×98
	No. of Modes	2
	Value of Variation in Each Mode	97
	No. of Iterations (For Optimization)	02
	Size of the Principal Component Matrix - Raw Dataset)	68×94
	Size of the Principal Component Matrix - DOSITS)	64×84

As mentioned earlier, spectrograms have been used as inputs to both the said techniques. Moreover, simulations have been done in two different manners, first the raw spectrograms of the acoustic samples were fed to the classifier, i.e. convolutional neural network (CNN), for classification whereas in second case multi-linear principal component analysis (MPCA) was first used for feature learning/ dimensionality reduction and then, the resultant feature set was fed as input to the classifier i.e. convolutional neural network (CNN). Table 4.14 and 4.15 illustrates the classification results for both the data sets i.e. synthetic and ship data, using the said techniques. Moreover, from simulation results it can be deduced that compare to the usage of raw spectrograms, usage of feature set learned from multi-linear principal component analysis (MPCA) gave better classification accuracies with a minimum of 33% reduction in overall dimensions of the feature set. Moreover, CNN with raw spectrograms as inputs gave very good classifica-

tion accuracies compared to results obtained using other approaches but due to the size of the feature vector and spareness of features in the spectrogram, the classifier missed some of the fine details present in the spectrum. Further, usage of MPCA improved the classification results with reduction in size of the feature vector, thus, enabling the classifier to learn better.

Table 4.14: DOSITS - Using CNN and MPCA - %Recognition - Feature Extraction Techniques vs Classification Techniques

			Classification Technique
		SNR	Convolutional Neural Network (CNN)
Feature Extract. Techni.	**Raw Input (Spectrogram)**	0	95.11%
		10	97.47%
		20	98.23%
	Multi-linear Principal Component Analysis (MPCA)	0	98.2%
		10	98.9%
		20	99.3%

Table 4.15: Raw Dataset - Using CNN and MPCA - %Recognition - Feature Extraction Techniques vs Classification Techniques

			Classification Technique
		SNR	Convolutional Neural Network (CNN)
Feature Extract. Techni.	**Raw Input (Spectrogram)**	0	96.5%
		10	98.21%
		20	98.31%
	Multi-linear Principal Component Analysis (MPCA)	0	98.9%
		10	99.1%
		20	99.4%

Multi-linear principal component analysis (MPCA) with convolution neural network (CNN) have produced best classification accuracies compared to other standard & renowned feature extractors and classifiers used in this study. In addition, from table 4.3 and 4.4, it can also be concluded that amongst classifiers, dynamic time warping (DTW) has performed better than all other classifiers except CNN. However, it is computationally more expensive and takes more processing time during testing phase as compared to other classification schemes. Moreover, with other classifiers the recognition rates have increased significantly with the increase in signal-to-noise ratio (SNR) of the feature set. Further, among the cepstral analysis techniques, linear predictive cepstral coefficient (LPCC) produced the most robust and discriminating feature sets along with mel-frequency cepstral coefficient (MFCC) and gammatone cepstral coefficient (GTCC) whose classification accuracies were quite close to GTCC.

5 Conclusion

Passive sonar signal detection and classification system aims to detect and classify targets coming from different directions. It has major applications in military and defence settings. Moreover, signal detection and classification is difficult and a challenge due to the environment's non-stationary nature. Aim of this study was to present a comprehensive study for detection and classification of underwater objects based on signals acquired via sonar platform. Performance of the system was evaluated using two sets of data i.e. a raw dataset acquired via sonar platform, having samples belonging to 4 distinct classes of ships and a synthetic dataset downloaded from DOSITS [79], having samples belonging to 20 distinct classes of underwater objects i.e. sea species and man-made objects. The front-end unit used for signal detection, comprised of major acoustic signal analysis techniques, including, wavelet analysis, renowned sonar signal detection techniques i.e. detection of envelope modulation on noise (DEMON) and low frequency analysis & ranging (LOFAR) and some of the most acknowledged speech signal processing techniques i.e. linear predictive analysis (LPA), linear predictive cepstral coefficient (LPCC), perceptual linear prediction (PLP/ BFCC), mel-frequency cepstral coefficient (MFCC) and gammatone cepstral coefficient (GTCC). Moreover, the back-end unit for signal classification used discriminating feature set obtained using aforementioned detection techniques. For classification, machine learning techniques have been employed i.e. variants of neural networks and template matching. System utilized five classifiers i.e. multilayer feed-forward neural network (MFNN), variable learning rate feed-forward neural network (VLR-NN), radial-basis function neural network (RBF-NN), convolutional neural network (CNN) and dynamic time warping (DTW). Moreover, effects of dimensionality reduction on classification rates were also observed. For dimensionality reduction, Multi-linear principal component analysis (MPCA), principal component analysis (PCA) and linear discriminant analysis (LDA) were used. The system was tested under noisy conditions, that is, under the effects of additive white Gaussian noise (AWGN) at different levels of signal-to-noise Ratio (SNR) i.e. from -20 dB to 20 dB.

Results obtained for both the datasets i.e. raw and synthetic, have made similar reading and based on simulation results following conclusions are in order,

- Amongst spectral feature sets, LOFAR features performed better than DEMON features in terms of classification accuracies.
- Amongst cepstral analysis techniques, LPCC features outperformed other approaches in terms of classification accuracies.
- Amongst cepstral analysis techniques, MFCC and GTCC were second and third best after LPCC in terms of classification accuracies, respectively.
- In speech recognition paradigm, GTCC has outperformed every other cepstral analysis technique in terms of classification accuracy, specially in noisy conditions but as per the simulation results it didn't work well with underwater acoustic sequences.
- Feature set acquired using LPA didn't perform well in terms of classification accuracies compared to other feature sets obtained using spectral and cepstral approaches.
- Overall, cepstral features gave better classification accuracies compared to pure spectral features.
- Classification accuracy significantly increased with the increase in signal-to-noise ratio.
- At low SNR, LPA and LPCC features gave better classification accuracies.
- Wavelet features didn't perform well compared to cepstral and spectral features, specially at low SNR.
- Amongst all classifiers, CNN, a deep learning approach, produced best classification results. Moreover, CNN was fed with a spectrogram (2D or TF Matrix) as an input compared to other classification approaches where 1D feature vector was used as input to the classifier.

- Apart from CNN, DTW produced better classification results. Moreover in terms of classification accuracies, DTW lagged behind CNN at low values of SNR.

- Between RBF, MFNN and VLR-NN, comparatively RBF has performed way better in terms of classification accuracies.

- Amongst all classifiers, DTW took most processing time during testing phase.

- Linear subspace learning (LSL) approaches, namely, principal component analysis (PCA) and linear discriminant analysis (LDA), and Multi-linear subspace learning (MSL) approach, multi-linear principal component analysis (MPCA) have been used for dimensionality reduction. Moreover, LSL techniques were used to reduce dimension of 1D feature vector whereas MSL approach was used for reducing dimensions of a 2D vector.

- For dimensionality reduction of spectral features MSL and LSL approaches were used. Moreover, a spectrogram was used as an input to the MSL approach whereas LOFAR feature set was used as input to the LSL approaches.

- Amongst both MSL and LSL techniques, transformed feature set obtained using MSL technique gave best results in terms of classification accuracies.

- Amongst the LSL techniques, principal component analysis (PCA) performed relatively better compared to linear discriminant analysis (LDA) in terms of classification accuracy.

- Overall, amongst all outlined approaches, combination of MPCA and CNN gave best classification results with an accuracy up to 99.4%.

6 Future Work

There are numerous problems and challenges still exist that halts the development of a proper framework offering sustained performance levels in object detection and classification in underwater environment. These challenges open up a lot of research areas and give direction that needs to be pursued in order to have a proper framework for signal detection and classification in underwater environment.

One such area is to employ techniques resulting in optimum discriminating feature set and to complement the feature extraction process a strong classifier is to be selected to have better recognition rates. Signal transformation, signal analysis and machine learning methods should be employed to obtain optimal results for the temporally varying signalling sources. Another aspect of sonar signal processing is mathematical modelling of underwater environment and noise producing sources. Good modelling can have significant effect in obtaining overall good classification rates. It is another area which can be explored and worked at to develop an efficient sonar signal processing unit. Machine learning techniques i.e. Deep Learning, prediction methods and statistical inference models i.e. Markov models and it's variants, can be used to model the underwater objects. Another area in this field pertains to source separation. It is pivotal in attaining good detection and classification results. For efficient detection and classification, source separation is of equal importance, as important as to have a good feature extractor and a strong classifier. Techniques, including independent component analysis, blind source separation and its variants are vastly employed to achieve good source separation. Several algorithms for independent component analysis (ICA) exist that works well in noisy environments where source separation is difficult, namely, fastICA, joint approximate diagonalization of eigen matrices (JADE), multiplicative newton-like algorithm and time domain-blind source separation etc.

Another area in this domain is to devise and employ methods to develop computa-

tionally low-cost systems, such that, it should not effect system's performance levels.

Work presented in this study can be expanded to different signal processing schemes, larger datasets and to sound samples acquired in different conditions and environments. The presented study can also be expanded for detailed analysis of each approach for performance evaluation at micro scale. Moreover, optimum signal processing and machine learning techniques must be employed to improve the overall recognition rates.

Finally, in terms of application, the most prominent areas are commercial and military settings, that is, sea tomography, shoal fish detection, surveillance of coastal areas and many more. Perhaps, this framework with little variations can be employed in applications where acoustic signal detection and prediction is required.

BIBLIOGRAPHY

Bibliography

[1] Ossama Abdel-Hamid, Abdel-Rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, 22(10):1533–1545, 2014.

[2] Waleed H. Abdulla. Auditory based feature vectors for speech recognition systems. In *In Advance in Communication and Software Technologies, N.E. Mastorakis & V.V. Kluev, Editor, WSEAS*, pages 231–236. Press.

[3] T. Aboulnasr and K. Mayyas. A robust variable step-size lms-type algorithm: analysis and simulations. *Signal Processing, IEEE Transactions on*, 45(3):631–639, Mar 1997.

[4] Paolo Antognetti and Velijko Milutinovic. *Neural networks: concepts, applications, and implementations.* Prentice Hall Press, 1991.

[5] Daniel Avitzour. A maximum likelihood approach to data association. *IEEE Transactions on Aerospace and Electronic Systems*, 28(2):560–566, 1992.

[6] Mohtashim Baqar, Sohaib Azhar, Zeeshan Iqbal, Irfan Shakeel, Laeeq

Ahmed, and Muhammad Moinuddin. Efficient iris recognition system based on dual boundary detection using robust variable learning rate multilayer feed forward neural network. In *Information Assurance and Security (IAS), 2011 7th International Conference on*, pages 326–330. IEEE, 2011.

[7] Erwin H Bareiss. Numerical solution of linear equations with toeplitz and vector toeplitz matrices. *Numerische Mathematik*, 13(5):404–424, 1969.

[8] Nicolas Basalto, Roberto Bellotti, Francesco De Carlo, Paolo Facchi, Ester Pantaleo, and Saverio Pascazio. Hausdorff clustering of financial time series. *Physica A: Statistical Mechanics and its Applications*, 379(2):635–644, 2007.

[9] Claudio Becchetti and Lucio Prina Ricotti. Speech recognition–theory and c++ implementation, 1999.

[10] Herbert Buchner, Robert Aichner, and Walter Kellermann. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. *Speech and Audio Processing, IEEE Transactions on*, 13(1):120–134, 2005.

[11] William S Burdic. *Underwater acoustic system analysis*. Prentice Hall, 1991.

[12] David J Burr. Experiments on neural net recognition of spoken and written text. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(7):1162–1168, 1988.

[13] David A Castañón. New assignment algorithms for data association.

In *Aerospace Sensing*, pages 313–323. International Society for Optics and Photonics, 1992.

[14] Sheng Chen, Colin FN Cowan, and Peter M Grant. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2):302–309, 1991.

[15] Ming-Chi Lin Chin-Hsing Chen, Jiann-Der Lee. Classification of underwater signals using neural networks. *Tamkang Journal of Science and Engineering*, Vol. 3, No.1:31–48, 2000.

[16] DW Cottle and DJ Hamilton. All neural network sonar discrimination system. In *Neural Networks for Ocean Engineering, 1991., IEEE Conference on*, pages 13–19. IEEE, 1991.

[17] Kane A Cunningham and David C Mountain. Simulated masking of right whale sounds by shipping noise: Incorporating a model of the auditory periphery. *The Journal of the Acoustical Society of America*, 135(3):1632–1640, 2014.

[18] Steven B Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(4):357–366, 1980.

[19] NN De Moura, JM De Seixas, and Ricardo Ramos. *Passive Sonar Signal Detection and Classification Based on Independent Component Analysis*. Citeseer, 2011.

[20] Jose Manoel de Seixas, NN De Moura, et al. Preprocessing passive sonar signals for neural classification. *IET radar, sonar & navigation*,

5(6):605–612, 2011.

[21] Li Deng, Jinyu Li, Jui-Ting Huang, Kaisheng Yao, Dong Yu, Frank Seide, Michael Seltzer, Geoff Zweig, Xiaodong He, Jason Williams, et al. Recent advances in deep learning for speech research at microsoft. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8604–8608. IEEE, 2013.

[22] JC Di Martino, J-P Haton, and A Laporte. Lofargram line tracking by multistage decision process. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 1, pages 317–320. IEEE, 1993.

[23] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification (pt. 1). 2000.

[24] Donald P Eickstedt and Henrik Schmidt. A low-frequency sonar for sensor-adaptive, multistatic, detection and classification of underwater targets with auvs. In *OCEANS 2003. Proceedings*, volume 3, pages 1440–1447. IEEE, 2003.

[25] M Farrokhrooz and M Karimi. Ship noise classification using probabilistic neural network and ar model coefficients. In *Oceans 2005-Europe*, volume 2, pages 1107–1110. IEEE, 2005.

[26] Mehdi Farrokhrooz and Mahmood Karimi. Marine vessels acoustic radiated noise classification in passive sonar using probabilistic neural network and spectral features. *Intelligent Automation & Soft Computing*, 17(3):369–383, 2011.

[27] Hervé Gauvrit, J-P Le Cadre, and Claude Jauffret. A formulation of

multitarget tracking as an incomplete data problem. *Aerospace and Electronic Systems, IEEE Transactions on*, 33(4):1242–1257, 1997.

[28] Joydeep Ghosh, Larry M Deuser, and Steven D Beck. A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals. *Oceanic Engineering, IEEE Journal of*, 17(4):351–363, 1992.

[29] Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1):103–138, 1990.

[30] R Paul Gorman and Terrence J Sejnowski. Learned classification of sonar targets using a massively parallel network. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(7):1135–1140, 1988.

[31] Ronald L Greene and Robert L Field. Classification of underwater acoustic transients by artificial neural networks. In *Neural Networks for Ocean Engineering, 1991., IEEE Conference on*, pages 275–281. IEEE, 1991.

[32] Simon Haykin. Neural network - a comprehensive foundation. *Neural Networks*, 2(2004), 2004.

[33] Simon S Haykin, Simon S Haykin, Simon S Haykin, and Simon S Haykin. *Neural networks and learning machines*, volume 3. Pearson Education Upper Saddle River, 2009.

[34] H. Hermansky. Perceptual linear predictive (plp) analysis for speech. *The Journal of the Acoustical Society of America*, pages 1738–1752, 1990.

[35] Hynek Hermansky, Brian A Hanson, and Hisashi Wakita. Perceptually based linear predictive analysis of speech. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85.*, volume 10, pages 509–512. IEEE, 1985.

[36] Hynek Hermansky, Nelson Morgan, Aruna Bayya, and Phil Kohn. The challenge of inverse-e: the rasta-plp method. In *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*, pages 800–804. IEEE, 1991.

[37] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[38] Carine Hue, Jean-Pierre Le Cadre, and Patrick Pérez. Sequential monte carlo methods for multiple target tracking and data fusion. *Signal Processing, IEEE Transactions on*, 50(2):309–325, 2002.

[39] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[40] Nicholas Harold Klausner. Underwater target detection using multiple disparate sonar platforms. 2007.

[41] Amlan Kundu, George C Chen, and Charles E Persons. Transient sonar signal classification using hidden markov models and neural nets. *Oceanic Engineering, IEEE Journal of*, 19(1):87–99, 1994.

[42] Thomas A Lampert and Simon EM OKeefe. A survey of spectrogram track detection algorithms. *Applied acoustics*, 71(2):87–100, 2010.

[43] Qihu Li, Jinlin Wang, and Wei Wei. An application of expert system in recognition of radiated noise of underwater target. In *OCEANS'95. MTS/IEEE. Challenges of Our Changing Global Environment. Conference Proceedings.*, volume 1, pages 404–408. IEEE, 1995.

[44] JG Lourens. Classification of ships using underwater radiated noise. In *Communications and Signal Processing, 1988. Proceedings., COMSIG 88. Southern African Conference on*, pages 130–134. IEEE, 1988.

[45] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. Mpca: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39, 2008.

[46] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, April 1975.

[47] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[48] Charles L Morefield. Application of 0-1 integer programming to multitarget tracking problems. *Automatic Control, IEEE Transactions on*, 22(3):302–312, 1977.

[49] David P Morgan and Christopher L Scofield. *Neural networks and speech processing*. Springer, 1991.

[50] Jill K Nelson and Hossein Roufarshbaf. A tree search approach to target tracking in clutter. In *Information Fusion, 2009. FUSION'09. 12th International Conference on*, pages 834–841. IEEE, 2009.

[51] R. O. Nielsen. *Sonar signal processing*. Artech House, 1991.

[52] Zoran Salcic Octavian Cheng, Waleed Abdulla. Performance evaluation of front-end processing for speech recognition systems. Master's thesis, Electrical and Computer Engineering Department, School of Engineering, The University of Auckland, 2005.

[53] R. B. Palm. Prediction as a candidate for learning deep hierarchical models of data. Master's thesis, 2012.

[54] Krishna R Pattipati, Somnath Deb, Yaakov Bar-Shalom, and Robert B Washburn Jr. A new relaxation algorithm and passive sensor data association. *Automatic Control, IEEE Transactions on*, 37(2):198–213, 1992.

[55] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[56] Hossein Peyvandi, Hossein Roufarshbaf, Mehdi Farrokhrooz, and Sung-Joon Park. *SONAR systems and underwater signal processing: classic and modern approaches*. INTECH Open Access Publisher, 2011.

[57] Aubrey B Poore and Nenad Rijavec. Multitarget tracking and multidimensional assignment problems. In *Orlando'91, Orlando, FL*,

pages 345–356. International Society for Optics and Photonics, 1991.

[58] L. Rabiner and R. Schafer. *Digital Processing of Speech Signals*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1978.

[59] Lawrence Rabiner and Biing-Hwang Juang. Fundamentals of speech recognition. 1993.

[60] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[61] R Rajagopal, B Sankaranarayanan, and P Rao. Target classification in a passive sonar-an expert system approach. In *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, pages 2911–2914. IEEE, 1990.

[62] Donald B Reid. An algorithm for tracking multiple targets. *Automatic Control, IEEE Transactions on*, 24(6):843–854, 1979.

[63] Donald Ross. Noise sources, radiation and mitigation. In *Underwater Acoustics and Signal Processing*, pages 3–28. Springer, 1981.

[64] Donald Ross. *Mechanics of underwater noise*. Elsevier, 2013.

[65] Hossein Roufarshbaf. *A tree search approach to detection and estimation with application to communications and tracking*. PhD thesis, George Mason University, 2011.

[66] Hossein Roufarshbaf and Jill K Nelson. Target tracking via a sampling stack-based approach. In *Signals, Systems and Computers,*

2009 Conference Record of the Forty-Third Asilomar Conference on, pages 1327–1331. IEEE, 2009.

[67] Hossein Roufarshbaf and Jill K Nelson. Evaluation of multistatic tree-search based tracking on the seabar dataset. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–8. IEEE, 2010.

[68] David E Rumelhart, James L McClelland, PDP Research Group, et al. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1-2. *Cambridge, MA*, 1986.

[69] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 26(1):43–49, 1978.

[70] Bernhard Scholkopft and Klaus-Robert Mullert. Fisher discriminant analysis with kernels. *Neural networks for signal processing IX*, 1(1):1, 1999.

[71] K Sam Shanmugan and Arthur M Breipohl. Random signals: detection, estimation, and data analysis. 1988.

[72] Robert Singer and John Stein. An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems. In *1971 IEEE Conference on Decision and Control*, number 10, pages 171–175, 1971.

[73] M. Slaney. An efficient implementation of the patterson-holdsworth auditory filter bank. Technical report, Apple Technical Report No. 35, Advanced Technology Group, Apple Computer, Inc., Cupertino, CA,, 1993.

[74] William Soares-Filho, José Manoel De Seixas, and Luiz Pereira Calôba. Averaging spectra to improve the classification of the noise radiated by ships using neural networks. In *Neural Networks, 2000. Proceedings. Sixth Brazilian Symposium on*, pages 156–161. IEEE, 2000.

[75] William Soares-Filho, Jose Manoel De Seixas, and L Pereira Caloba. Principal component analysis for classifying passive sonar signals. In *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, volume 3, pages 592–595. IEEE, 2001.

[76] Stanley Smith Stevens, John Volkmann, and Edwin B Newman. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3):185–190, 1937.

[77] Roy L Streit and Tod E Luginbuhl. A probabilistic multi-hypothesis tracking algorithm without enumeration and pruning. In *Proc. 6th Joint Sevice Data Fusion Symp*, volume 16, 1993.

[78] Roy L Streit and Tod E Luginbuhl. Maximum likelihood method for probabilistic multihypothesis tracking. In *SPIE's International Symposium on Optical Engineering and Photonics in Aerospace Sensing*, pages 394–405. International Society for Optics and Photonics, 1994.

[79] Graduate School of Oceanography University of Rhode Island. Discovery of sound in the sea.

[80] Robert J Urick. *Principles of underwater sound for engineers*. Tata McGraw-Hill Education, 1967.

[81] Michael K Ward and Maryhelen Stevenson. Sonar signal detection

and classification using artificial neural networks. In *Electrical and Computer Engineering, 2000 Canadian Conference on*, volume 2, pages 717–721. IEEE, 2000.

[82] P.D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio Electroacoustics*, AU-15:7073, 1967.

[83] Gordon M Wenz. Acoustic ambient noise in the ocean: spectra and sources. *The Journal of the Acoustical Society of America*, 34(12):1936–1956, 1962.

[84] He Xi-Ying, Cheng Jin-Fang, He Guang-Jin, and Li Nan. Application of bp neural network and higher order spectrum for ship-radiated noise classification. In *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, volume 1, pages V1–712. IEEE, 2010.