Radar Target Micro-Doppler Signature Classification

Graeme Edward Smith

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Electronic and Electrical Engineering

University College London

2008

I, Graeme Edward Smith, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis reports on research into the field of Micro-Doppler Signature (μ-DS) based radar Automatic Target Recognition (ATR) with additional contributions to general radar ATR methodology. The μ-DS based part of the research contributes to three distinct areas: time domain classification; frequency domain classification; and multiperspective μ-DS classification that includes the development of a theory for the multistatic μ-DS. The contribution to general radar ATR is the proposal of a methodology to allow better evaluation of potential approaches and to allow comparison between different studies.

The proposed methodology is based around a "black box" model of a radar ATR system that, critically, includes a threshold to detect inputs that are previously unknown to the system. From this model a set of five evaluation metrics are defined. The metrics increase the understanding of the classifier's performance from the common probability of correct classification, that reports how often the classifier correctly identifies an input, to understanding how reliable it is, how capable it is of generalizing from the reference data, and how effective its unknown input detection is. Additionally, the significance of performance prediction is discussed and a preliminary method to estimate how well a classifier should perform is developed. The proposed methodology is then used to evaluate the μ-DS based radar ATR approaches considered.

The time domain classification investigation is based around using Dynamic Time Warping (DTW) to identify radar targets based on their μ-DS. DTW is a speech processing technique that classifies data series by comparing them with a pre-classified reference dataset. This is comparable to the common k-Nearest Neighbour (k-NN) algorithm, so k-NN is used as a benchmark against which to evaluate DTW's performance. The DTW approach is observed to work well. It achieved high probability of correct classification and reliability as well as being able to detect inputs of unknown class. However, the classifier's ability to generalize from the reference data is less impressive and it performed only slightly better than a random selection from the possible output classes. Difficulties in classifying the μ-DS in the time domain are identified from the k-NN results prompting a change to the frequency domain.

Processing the μ-DS in the frequency domain permitted the development of an advanced feature extraction routine to maximize the separation of the target classes and therefore reduce the effort required to classify them. The frequency domain also permitted the use of the performance prediction method developed as part of the radar ATR methodology and the introduction of a naïve Bayesian approach to classification. The results for the DTW and k-NN classifiers in the frequency domain were comparable to the time domain, an unexpected result since it was anticipated that the μ-DS would be easier to

classify in the frequency domain. However, the naïve Bayesian classifier produced excellent results that matched with the predicted performance suggesting it could not be bettered. With a successful classifier, that would be suitable for real-world use, developed attention turned to the possibilities offered by the multistatic μ-DS.

Multiperspective radar ATR uses data collected from different target aspects simultaneously to improve classification rates. It has been demonstrated successful for some of the alternatives to μ-DS based ATR and it was therefore speculated that it might improve the performance of μ-DS ATR solutions. The multiple perspectives required for the classifier were gathered using a multistatic radar developed at University College London (UCL). The production of a dataset, and its subsequent analysis, resulted in the first reported findings in the novel field of the multistatic μ-DS theory. Unfortunately, the nature of the radar used resulted in limited micro-Doppler being observed in the collected data and this reduced its value for classification testing. An attempt to use DTW to perform multiperspective μ-DS ATR was made but the results were inconclusive. However, consideration of the improvements offered by multiperspective processing in alternative forms of ATR mean it is still expected that μ-DS based ATR would benefit from this processing.

Acknowledgements

The work contained in this thesis would not have been possible without the support of several people. I would like to take this opportunity to thank them.

At University College London I am indebted to Dr Karl Woodbridge and Prof. Chris Baker for the excellent supervision they have provided. I am also grateful for Shaun Doughty's incredible patience when listening to my crack-pot ideas. No matter how animated I became Shaun was always there to calmly draw a phasor diagram on our whiteboard and help convert my rambling theories into solid engineering. Rounding off my UCL support team is Michele Vespe. Every coffee break we took generated enough ideas for a whole lifetime of research, let alone one PhD.

From industry I would like to thank Jeremy Langley, Steve Roberts, Bernard Walsh and Tony Kinghorn from Selex Sensors and Airborne Systems. Their belief in my abilities, both when I worked with them and while I have been studying at UCL, has always been a great source of motivation to me. I am also grateful for the support offered by Andy Stove and Simon Watts of Thales. The data they supplied and our discussions about my research were crucial in keeping me firmly grounded in practical engineering rather than abstract theory.

To close, I want to acknowledge the immense support I have received from my family. I dedicate the work presented here to them.

Contents

List of Figures

List of Tables

Glossary Of Terms

ADC Analogue to Digital Converter

AM Amplitude Modulation

ANN Artificial Neural Network

ATR Automatic Target Recognition

BW Bandwidth

CW Continuous Wave

dB Decibel

DC Direct Current

DFT Discrete Fourier Transform

DSP Digital Signal Processing

DTW Dynamic Time Warping

FFT Fast Fourier Transform

FM Frequency Modulation

GD Global Distance

GMM Gaussian Mixture Model

HMM Hidden Markov Model

HRR High Range Resolution

HRRP High Range Resolution Profile

IF Intermediate Frequency

IFF Identify Friend or Foe

ISAR Inverse Synthetic Aperture Radar

JEM Jet Engine Modulation

k-NN k-Nearest Neighbour

LOS Line Of Sight

MDL Main Doppler Line

MSTAR Man-portable Surveillance and Target Acquisition Radar

MTI Moving Target Indication

PCA Principal Component Analysis

PRF Pulse Repetition Frequency

PRI Pulse Repetition Interval

PW Pulse Width

NGD Normalized Global Distance

RCS Radar Cross Section

RF Radio Frequency

ROC Receiver Operating Characteristic

SAR Synthetic Aperture Radar

SNR Signal to Noise Ratio

SSR Secondary Surveillance Radar

STFT Short Time Fourier Transform

UCL University College London

μ**-DS** Micro-Doppler Signature

Chapter 1

Introduction

The background to the research is introduced at a high level and placed in an historical context. An abridged history of the development of radar is presented and it is noted that throughout this history attempts have been made to extend the capabilities of radar to include target recognition. As a close to the research background, the concept of the Micro-Doppler Signature (μ-DS) is introduced and its value to target recognition discussed.

The contributions of the research to the discipline of radar target recognition are then listed. The chapter ends with a description of the thesis structure.

1.1 Overview

The traditional function of radar is to detect targets and indicate their range, hence *radar* from *ra*dio *d*etection *a*nd *r*anging. The first radars were developed in the 1930s although the underlying principle was proved almost fifty years before. Toward the end of the 1880s Henrich Hertz conducted a series of experiments to demonstrate that electromagnetic waves could be reflected from metal objects (Britannica, 2007). Following these experiments in 1904 another German, Christian Hülsmeyer, filed a patent for a ship detection device, intended to aid sea navigation in fog, that relied on the principles demonstrated by Hertz. However, Hülsmeyer's initial device was unable to indicate range and it never became popular, even after a later improvement provided range information (Swords, 1986). During the First World War mainland Britain faced a threat of bombardment by early German aircraft that required long range detection to counter. Rather than use active electromagnetic systems the British armed forces developed a form of passive acoustic detection that relied on large, fixed, concrete 'sound mirrors' to listen for the enemy aircraft (Scarth, 1999). It was not until after the First World War, with the potential threat of attack by the Nazi's Luftwaffe, that Britain began the development of radar. In January 1935 the Director of Scientific Research for British Air Defence approached Robert Watson-Watt, of the National Physics Laboratory, about the possibility of building a 'death ray' using electromagnetic waves that could disable aircraft. In the following two months Watson-Watt produced two memoranda: the first

explaining why it was impractical to build the 'death ray' and the second detailing how electromagnetic waves could be used to detect aircraft at long ranges. These lead to a series of experiments culminating in the development of the Chain Home radar system that is regarded as being an essential element of the Royal Air Force's victory in the Battle of Britain (Swords, 1986). Research into improving and refining radar was conducted throughout the Second World War and continues to this day with radar systems becoming smaller, more powerful and gaining functionality all the time.

In the period following the Second World War there were many advances in radar theory, although it was not always possible to implement the ideas immediately due to hardware limitations. During the war itself pulse delay systems had been used to create Moving Target Indication (MTI) radar and with the development of more stable oscillators it became possible to develop full pulse-Doppler radars, that could measure target velocity in addition to detecting motion, through the late 1950 and 1960s (Perkins et al., 1984). Constraints on the cost and size of the Radio Frequency (RF) filters required, however, meant that the dominance of the pulse-Doppler approach did not come until the advent of low cost digital processing (Stimpson, 1998). Also in development during the 1950s and 60s was the phased array antenna (Brookner, 2006) in which a series of small transmitting elements operate collectively to enable dynamic formation of the antenna beam pattern. With the digital processing revolution of the 1970s Synthetic Aperture Radar (SAR) technology allowed the gathering of high resolution ground images by radar (Stimpson, 1998). This technology had been available previously, although without digital processing intricate optical methods had to be employed and performance was limited. More recently, the last ten to fifteen years have seen the emergence of affordable high-frequency RF components permitting the development of millimetre-wave (77 and 94 GHz) radar. These devices have limited range, but a high bandwidth permitting fine range resolution and are becoming common in applications as diverse as airport surface monitoring (Ferri et al., 2003) and automotive radars in top of the range cars (Wenger, 2005). With all these developments it is now common to find radar in situations as diverse as: air surveillance for both civil and military operations; weather monitoring and prediction; remote sensing; mine and underground feature detection; Earth surface monitoring; medical imaging; and traffic speed enforcement. Yet despite these advances there is still no standard method for performing target recognition using radar alone.

Attempts to add Automatic Target Recognition (ATR) to radar's functionality have been made throughout its development. As early as 1937 experiments were conducted in which resonant dipoles were added to friendly aircraft so that their returns were distinctive from those of hostile aircraft (Swords, 1986). It was realized that such a system would have limited use when several aircraft flew in formation and focus shifted to using Secondary Surveillance Radar (SSR). By placing a transponder on targets to be observed by the radar they become 'co-operative targets' that transmit an enhanced version of the radar signal when illuminated that contains encoded information about their identity. Similar technology

is used in both civilian and military applications, although in military circles the technology is referred to as Identify Friend or Foe (IFF), and both suffer from the weakness that they require the target to provide a truthful identity encoding. It is therefore preferable that a radar be able to determine the type or class of its target based solely on the properties of the backscattered signal and without reliance on external inputs such as those from the SSR or a human operator. Such an extension to radar's capability requires a substantial increase in functionality. The detection problem requires solely that the radar be able to determine that a backscatter signal is present. The problem of target recognition, however, requires that following detection the backscatter signal be analysed to extract detailed information about the target before performing classification of this information. It is this analysis and classification functionality that is generally referred to with the expression 'radar ATR'.

In his often cited *Introduction To Radar Systems, 2nd Ed.* Merril Skolnik lists nine radar principles and related phenomena that may be utilized by a radar to permit ATR (Skolnik, 1980). As is usual for this corner stone of radar literature Skolnik's list is comprehensive, despite being twenty-seven years old, and includes the most common approaches of: High Range Resolution (HRR) in which a one-dimensional image of the target is produced and classified; Jet Engine Modulation (JEM) in which the characteristic frequency modulations induced in the echo signal by the jet engines are identified; Radar Cross Section (RCS) fluctuation in which the angular variation of target's reflectivity is used as a discriminating feature; SAR whereby the motion of the radar is used to synthesize a large aperture antenna permitting the formation of a detailed terrain image in which targets can be recognized; and Inverse Synthetic Aperture Radar (ISAR) which is similar to SAR except that the target's motion is used to synthesize the large antenna aperture. Despite these approaches now being well understood they are still areas of active research since practical implementation—for ATR uses—is challenging. A particular example of this continued research is the transformation of JEM into the more general Micro-Doppler Signature (μ-DS).

The μ-DS of a target is an intricate frequency modulation of the radar echo signal that results from the superposition of the returns from the various moving components making up the target. A moving component may be a wheel, propeller, jet engine turbine blade, limb or even the target chassis if it is bouncing on suspension or vibrating with the engine (Chen et al., 2006; Schneider, 1987). Originally the μ-DS was only observed in targets including propellers or jet engines since these components moved rapidly and induced Doppler shifts large enough to be distinguished from the main target Doppler frequency (Schneider, 1987; Martin and Mulgrew, 1990, 1992). As the dynamic range of radars has increased the transition has been made from JEM to the more general μ-DS since many moving components may now induce detectable frequency modulation. Exploitation of the μ-DS for target recognition is common in battlefield radar where the operator may listen to an audio interpretation of the radar's echo signal (Stove and Sykes, 2003; Bilik et al., 2006). With training the operators become skilled at

identifying the signatures of different targets and are able to perform target classification. Automation of this process is desirable since relying on human operators can have several drawbacks: new operators must be trained; the operator may make errors when tired or working under stressful conditions; and the operator requires several seconds to perform an identification causing scanning to stop. Some progress has been made in automating μ-DS based radar ATR and the Thales MSTAR radar now contains a classifier that can distinguish between several target classes encountered on the battle field (Stove and Sykes, 2003).

1.2 Contribution Of This Study

The following broad areas have been contributed to by the research described here.

1.2.1 Advances On Using The μ-DS For Radar ATR

The principal focus of this thesis is to advance the use the the μ-DS in radar ATR; it is in this area that the main contributions have been made. The most significant contribution is development of a high performance radar ATR solution in Chapter 7. The solution is based on the naïve Bayesian classifier, uses the feature extraction procedure of Chapter 5 and conform to the "black box" model of Chapter 3. The performance of the classifier is directly comparable to a commercial system that is discussed in section 2.5, the literature survey. Despite its comparison with an existing system the naïve Bayesian classifier represents an advance: its classification is based purely on the μ-DS, the commercial system uses some additional system dependant inputs; and its conformance with the methodology of Chapter 3 allows a more robust analysis of its workings. The development of this high performance classifier was supported by advances made using template based classification techniques, specifically Dynamic Time Warping (DTW), in Chapter 6 and at the start of Chapter 7.

The primary contribution from the template classification approaches was the application of DTW to the problem of μ-DS based radar ATR. DTW is a technique from speech recognition that allows a comparison between two signals and is able to allow for misalignment of features within the signals (Sakoe and Chiba, 1978; Myers et al., 1980; Sankoff and Kruskal, 1999). The μ-DS theory that is described in Chapter 4 indicated that feature misalignment would be present in radar μ-DS data. In Chapter 6 and Chapter 7 DTW is shown capable of classifying μ-DS data although its ability to extrapolate (or generalize) from its reference data is limited when compared with the naïve Bayesian classifier. A further contribution of the DTW investigation was to consider classifying the raw radar data rather than perform feature extraction.

Time domain μ-DS data was successfully classified, using DTW, in Chapter 6. This processing is particularly significant since it involved no feature extraction. Limiting the feature extraction performed prior to the classifier acts to minimize the computational overhead associated with adding ATR to a

radar's functionality.

1.2.2 Extension Of Theory To Include Multistatic μ-DS

A multistatic version of the μ-DS was investigated to help overcome the problem of self-occlusion, where the target chassis obstructs the components that generate the signature, that is experienced in the monostatic case (Bell and Grubbs, 1993). Conventional radar operates monostatically, where transmission and reception is performed at the same site (Skolnik, 1980; Stimpson, 1998); the majority of μ-DS research has been conducted using such data (Chen et al., 2006; Stove and Sykes, 2003; Bilik et al., 2006; Smith et al., 2006a; Jahangir et al., 2003b,a; Jiajin, 2005). A bistatic version of the μ-DS has been considered, (Johnsen et al., 2003), although this approach still relies on observing a single target aspect and, as such, is limited by self-occlusion. A multistatic radar is able to observe a target from several aspects angles simultaneously and is therefore more robust to self-occlusion. A theory of multistatic μ-DS was developed and data from experimental trials analysed to validate the theory Chapter 8.

The extension of the μ-DS theory and the field trials performed to gather data are important contribution of the presented work. The literature survey, see section 2.5, did not reveal any previous publications in this area suggesting that the work of Chapter 8 is particularly novel. The developed theory was used to identify information redundancy in the multistatic data that may simplify the data fusion requirements of multiperspective classification systems. An attempt was made to use the available multistatic μ-DS data in a multiperspective classifier. However, the results were of reduced value due to limitations identified in the dataset.

1.2.3 Proposal Of Quantitative Method For Dataset Creation

In radar ATR research there is commonly a quantity of available data that must be divided into a reference and test datasets to conduct the investigation. The selection of reference data is usually random (Bilik et al., 2006; Jahangir et al., 2003b,a). In this research a more quantitative approach is taken based on the μ-DS signal's correlation distance (the duration over which the signal can be considered correlated). By using a quantitative selection method the repeatability of the experiments was increased and it became possible to relate classifier performance to the correlation of the signal.

1.2.4 Proposal Of Radar ATR Evaluation Metrics

The radar classification community has yet to agree on a standard methodology for evaluating the performance of radar ATR systems. This can make comparison of different approaches difficult and can even lead to results that are unintentionally misleading. As key contribution of this work was the proposal of a set of evaluation metrics that provide a comprehensive analysis of ATR solutions and do not depend on the particular classifier employed, see Chapter 3. The evaluation metrics extend the understanding of classifiers' performances from the conventional probability of correct classification to include four

further measures: the first, reliability, describes the degree to which a classifier's output may be trusted; the second, generalization, details the classifiers ability to extrapolate from its training data; while the third and fourth new metrics, the probability of false alarm and the probability of declaration, measure the success of the ATR system's ability to reject inputs from unknown target classes. To support the use of extended performance metrics some additional investigation was undertaken into predicting classifier performance prior to implementation.

Prediction of the expected value of probability of correct classification was undertaken for that data used with the naïve Bayesian classifier. The predictions were made using the Bhattacharyya bound on the error. Despite the need to make a number of assumptions regarding the nature of the data the results were impressive with the performance of the classifier being accurately predicted. Again, this work represents a particularly novel contribution since the literature survey identified no existing publications for μ-DS based radar ATR performance prediction.

1.3 Thesis Outline

Chapter one has provided an introduction to the research, clarified the contributions of the research and outlined the thesis. An overview of the how the function of radar has expanded with time was given; this provides the historical context to the work. It was seen how advances in RF and electronic engineering allowed radar to achieve increasingly sophisticated functionality, made it possible to observe the μ-DS and to include target recognition in radar's function. Once the context of the research was established its contributions were detailed. The advances in using the μ-DS for classification; the application of DTW to radar; the new methods for developing reference and test datasets; the introduction of the multistatic μ-DS; and the use of multiperspective classification have been described.

The second chapter introduces basic radar theory. A block model is used to outline the various stages of a generic radar architecture. Following this the limits of radar detection are discussed and the "radar equation" introduced and explained. Substantial coverage is given to the way Doppler information is processed by radars since it is of great relevance to the research. The notions of range and Doppler are then combined through the ambiguity diagram to facilitate a brief review of the impact of the waveform on radar performance. The chapter is brought to a close with a survey of the current literature on multistatic radar, radar ATR and the μ-DS. Multistatic radar is a special class of radar system that employ a distributed sensor network to gather target data from multiple aspects simultaneously. Considering how these systems may contribute to the target recognition problem forms a sizeable part of the research presented here. Target recognition may be achieved in a number of ways: conventional RCS classification is discussed alongside contemporary approaches such as HRR imaging, SAR and ISAR. Particular focus is given to the existing methods of recognizing a target based on its μ-DS. The limitations to the established methods are discussed providing a rationale for why this research was required. The latter

part of the chapter reviews the methodology behind evaluating radar ATR systems and identifies that improvement is needed.

Chapter three focuses on how classification systems may be evaluated. A hypothetical "black box" classifier is presented and this is used to explore the classification concepts important in radar ATR. Why a radar ATR system should be able to declare "unknown" is considered and a technique to modify common classifiers to make this declaration is discussed. A set of evaluation parameters are defined and the use of Receiver Operating Characteristic (ROC) curves to balance the parameters and set internal thresholds is explained. Also discussed in this chapter is a method for estimating the performance of a classifier prior to its implementation. This is a highly novel area of research and is significant since without performance prediction it is not possible to understand how successful a classifier has been at discriminating between targets.

The theory behind the μ-DS is presented in chapter four. Many radar targets are composed of structures that exhibit their own limited motion in addition to the targets general motion. Since all parts of the target scatter the radar signal the moving components will induce their own Doppler shifts alongside that resulting from the gross target motion. These extra shifts create the μ-DS discussed in the chapter. A simple simulation is developed to assist in the explanation of the theory. The chapter is closed by analysing a real target μ-DS from data collected using a Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) radar.

A critical part of a radar ATR system is feature extraction which derives the feature vector to be classified from the raw radar data; the process by which feature vectors are created from the MSTAR data is discussed in the five chapter. Additionally, specific processing required for the data is presented. The removal of ground clutter by the CLEAN algorithm is considered. How the data is divided from long duration signals into realistic short frames is described. And the selection of these frames into test and reference datasets is outlined. Furthermore, the method by which two additional target classes were synthesized, for use in "unknown" detection testing, is discussed.

The ideas behind monoperspective ATR in the time domain and the results obtained are presented in the sixth chapter. The DTW and k-Nearest Neighbour (k-NN) classification algorithms are presented, and how they may be applied to the radar ATR problem described. It is demonstrated that the classifiers obtain comparable results when using intensity data as when I and Q samples are used. Following this the capabilities of the classifiers are evaluated under various conditions. The chapter closes with a discussion of limitations observed in the time domain classification approach.

The seventh chapter also concerns monoperspective ATR but in the frequency domain. The DTW and k-NN classifiers are used again and the naïve Baysian classifier is introduced since it can be used with frequency domain data. The nature of the feature vectors used with the naïve Baysian classifier permit the use of the performance prediction method from Chapter 3. While the results for the DTW and k-NN

show that improvements in their feature extraction process would be needed to operate on the frequency domain data the naïve Baysian results are impressive. It is also seen that the high performance is a good match with the predicted performance suggesting the classifier can not be significantly improved. Again, the chapter is closed with a discussion of the classifier limitations.

The penultimate chapter discusses the use of multiple target perspectives to assist radar ATR. Multiperspective classification has been successful in other radar ATR approaches, such as High Range Resolution Profile (HRRP) based methods, and it is speculated that it will also improve μ-DS based ATR. The chapter begins with a discussion of how the monostatic μ-DS may be considered limited and how several of these limitation may be overcome by using a multistatic approach—this approach would then necessitate multiperspective ATR techniques to be used. A brief review of the theory of multistatic radar is given before a novel theory of the multistatic μ-DS is developed. This theory is used to produce a simulation of a personnel target and predict the features of its multistatic μ-DS. A set of field trials used to gather data to validate the developed theory are described and their results analysed. A preliminary attempt at performing multiperspective μ-DS radar ATR using this data is reported, but the results are inconclusive due to limitations identified in the data.

The thesis is brought to a close in chapter nine where the conclusions are drawn. The chapter begins with a discussion of the novel contributions made in the research and these lead to a list of definite conclusions that may be drawn. The chapter ends with a series of recommendations for future work to be conducted in the field.

Chapter 2

Radar Basics

Before it is possible to discuss radar Automatic Target Recognition (ATR) it is necessary to develop a basic understanding of the science behind the radar itself. This chapter introduces the basic concepts of radar engineering through the use of a "block" model that highlights the key stages of a generic radar system. The model used has a few, slight, variations from the standard models shown in radar texts to support some of the more advanced concepts introduced in later chapters, primarily the idea of a micro-Doppler shift and bi- and multistatic radars. If anything, however, these variations serve to make the model, and accompany discussion, more general rather than more specific.

The chapter is closed with a survey of relevant, contemporary literature. Initial attention is given to the difference between monostatic and multistatic radar before moving on to radar based ATR. The current status of micro-Doppler research is outlined and the attempts at classifying the Micro-Doppler Signature (μ-DS) reviewed. Attention then moves to the implications of processing multiple views of the same target simultaneously. The survey is closed by considering the underlying methodology by which radar ATR work is conducted.

The fundamentals of radar theory have been well covered in the open literature and there are many excellent text books available discussing all manner of radar theory. This section is based on the following general texts: *Introduction To Radar Systems* by Merill I. Skolnik (Skolnik, 1980), *Introduction To Airborne Radar* by George W. Stimpson (Stimpson, 1998) and *Understanding Radar Systems* by Simon Kingsley and Shaun Quegan (Kingsley and Quegan, 1992). To prevent repeatedly citing the same source citations have only been included in sections 2.1 to 2.4 for sources other than those listed above or if the point is contentious in the general literature. For all other topics the reader should refer to any general radar text, particularly those listed above.

2.1 Block Model Of A Radar System

A radar operates by transmitting a Radio Frequency (RF) signal and receiving echoes of that signal that are backscattered by the target. The radar is able to determine the direction of the target from the orientation of the antenna(s) used during transmission and reception and, if a suitable signal is used, the target range from the time delay between transmission and reception. Depending on the type of radar it may also be possible to determine the target's radial velocity by analysing the Doppler shift of the backscatter signal. The radar needs to perform several actions in determining the stated target properties. A stage by stage approach will be taken to ease explanation.

A simplified radar design can be considered as comprising four sections, each of which can then be broken into smaller blocks: the transmitter channel; the receiver channel; common elements to both channels; and the processing stage. The latter stage is responsible for signal processing, tracking and displaying information to the operator. Fig. 2.1 is a block diagram showing the first three sections of a simplified radar. In this particular diagram the transmit and receive channels are shown as having their own antennas although it is perfectly possible for them to share an antenna if a duplexer is added to prevent the transmission being fed directly into the receiver. (The two antenna approach is taken here to facilitate a discussion of bistatic and multistatic radar in Chapter 8.) Before examining the transmit and receive channels the common elements are considered.

Common to both the receive and transmit channels are two RF oscillators. The coherent local oscillator, referred to as the *coho*, is responsible for providing the coherent reference signal to the synchronous detector for use in detecting phase changes in the received signal. The frequency of the coho is up-converted to transmission frequency by mixing it with the high frequency output of the stable local oscillator, known as the *stalo*. In many system diagrams the coho, stalo and upconverter are represented as a single entity called the *receiver-exciter* that provides three output signals: the transmission signal, a local oscillator signal for downconversion of the received signal, and a reference signal for use in phase detection. Under such an arrangement the transmit channel, which is discussed next, has fewer components.

The transmitter stage of the radar consists of three operations. First, the output of the stalo is used to upconvert the coho's output to transmission frequency. Upconversion is a combined process in which the two signals are mixed together and unwanted mixing artefacts, such as harmonics and sidebands, are filtered out. Second, the low power transmission signal is passed through a power amplifier. It is necessary that the transmitted signal be of as high a power as possible since the processes of travelling the distance to and from the target, and backscattering, results in a severe attenuation, see section 2.2, and the more power the received signal has the better the probability of detection. The third part of the channel is an output modulator. It is common for a radar to pulse its transmission to permit measurement

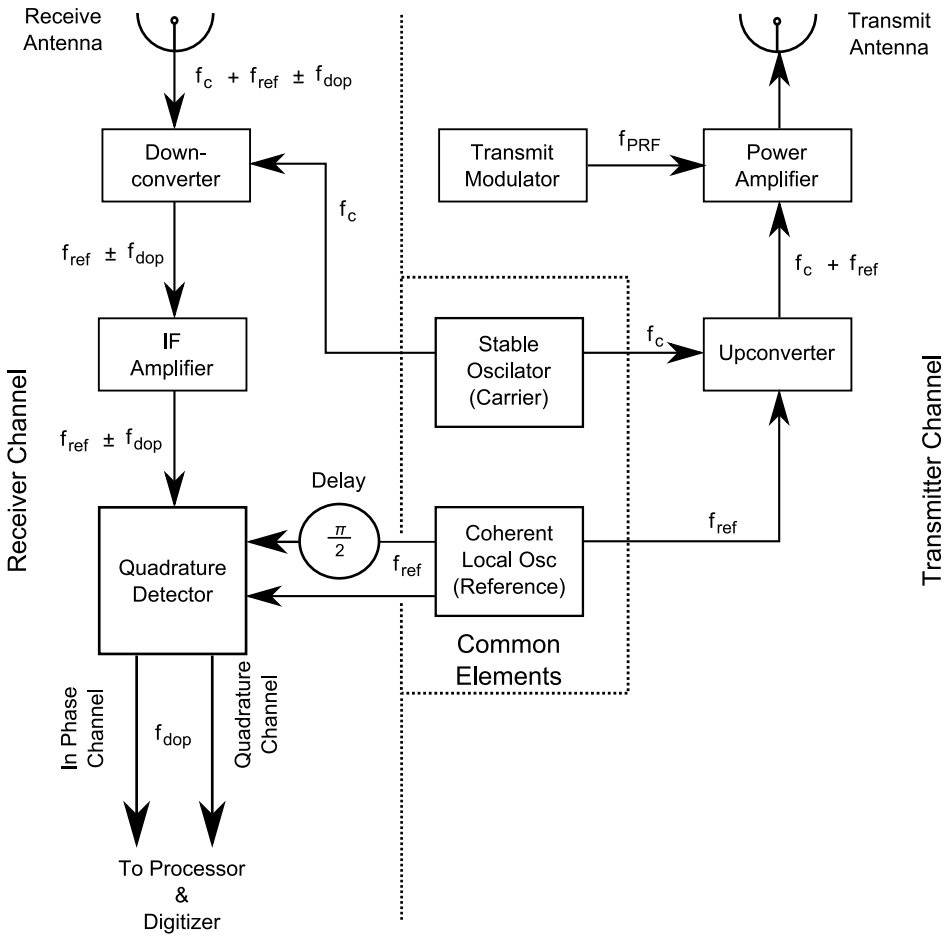


Figure 2.1: The basic components of a pulsed radar system.

of the target range. By pulsing the gain of the power amplifier an Amplitude Modulation (AM) is applied to the transmitted signal causing short pulses of RF energy to be emitted by the radar. The frequency with which these pulses are emitted is the radar's Pulse Repetition Frequency (PRF) and the duration of each pulse the Pulse Width (PW). Once the signal has been transmitted and backscattered it is ready to be received.

After a signal has been received by the radar it is down-converted, by mixing with the stalo output and filtering, to the radar Intermediate Frequency (IF). The IF is the same as the coho frequency and is the frequency at which the majority of the receiver components operate. Following downconversion the signal is amplified before passing to the quadrature detector. The quadrature detector serves two purposes: it applies a further downconversion to the IF signal to remove the coho signal, leaving only the target signal, and it splits the target signal into two channels one delayed by 90° in phase. It is the production of the delay channel that requires the second, 90° delayed, coho input. The two outputs are required to determine the sign of the Doppler shift.

Depending on the direction of target motion, towards or away from the radar, the Doppler shift will be either positive negative. Unfortunately, during the removal of the coho signal the sign of the Doppler shift is lost, see section 2.3, and only a positive frequency remains. Information about the sign of the Doppler shift can be kept by splitting the target signal into an in-phase (I) and quadrature (Q) channel. The I-channel is downconverted by the regular coho signal while the Q-channel is downconverted by the coho signal delayed by 90°of phase. Through comparison of the I- and Q-channels it is possible to determine the sign of the original Doppler shift, see section 2.3. Once the I- and Q-channels have been produced they pass on to the processor.

The fourth stage of the radar, the processor, can consist of many processes that depend upon the intended use for the radar. Broadly speaking these stages concern the application of signal processing techniques to the signal to allow the maximum amount of information about the target to be obtained. In older radar these stages may be accomplished through the intricate analogue methods while in modern systems the I- and Q- channel will be digitized as quickly as possible and Digital Signal Processing (DSP) applied. (Indeed, it is now common to digitize the signal before the quadrature detector and use digital techniques to produce the I- and Q-channel.) The processing stage will also control target tracking where an individual target is continuously monitored by the radar and the processor is required to control the antenna position to keep the target in the beamwidth. Once the data has been processed it is displayed to the operator or stored for further off-line processing.

2.2 Theoretical Limits Of Detection

2.2.1 The radar range equation

One of the principal functions of the radar is to determine the range to the target, Fig. 2.2. Assuming the

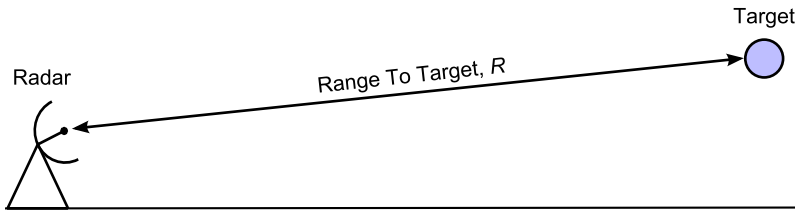


Figure 2.2: The range between the radar and its target.

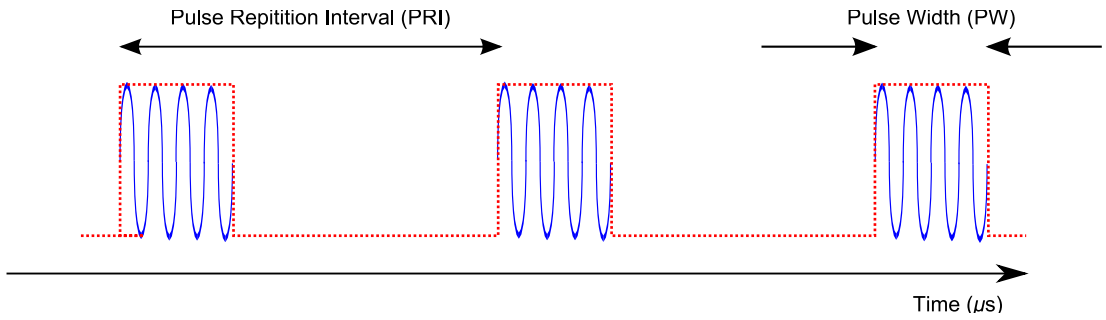


Figure 2.3: The structure of the pulse waveform.

radar uses a pulsed waveform, Fig. 2.3, the target range R may be derived from the time elapsed between pulse transmission and reception of the echo:

$$R = \frac{c\,\Delta t}{2} \tag{2.1}$$

where c is the speed of propagation, approximately 3×10^8 ms^{-1} for RF waves in the atmosphere, and Δt is the round trip time for the pulse. The duration between the pulses is referred to as the Pulse Repetition Interval (PRI) and the duration of the pulse the PW, Fig. 2.3. (The term PRI is interchangeable with PRF since they are the reciprocal of each other, i.e. $f_{\mathrm{PRF}} = 1/T_{\mathrm{PRI}}$.) The width of the pulse governs the range resolution of the radar, the minimum distance between two targets for which they are still separable, according to:

$$\Delta R = \frac{c\,\tau}{2} \tag{2.2}$$

where ΔR is the range resolution and τ the PW. Once the derivation of target range is understood the next logical concern is the maximum range of the radar.

The maximum range of the radar will depend on the power of the backscattered signal. If the radar transmits with a power P_{tx} then the received power is:

$$P_{\mathrm{rx}} = \frac{P_{\mathrm{tx}} G_{\mathrm{tx}} G_{\mathrm{rx}} \sigma \lambda^2 F^4}{(4\pi)^3 R^4 L}. \tag{2.3}$$

In (2.3) G_{tx} and G_{rx} are the gain of the transmitting and receiving antennas, σ the Radar Cross Section (RCS) of the target, λ is the carrier frequency wavelength, F^4 the pattern propagation factor, R the range

of the target and L the system losses. The gain of an antenna is a measure of its ability to amplify the radiation passing through it and to focus that radiation in a particular direction. It monostatic radar, the most common type of radar, the same antenna is used to transmit and receive the signal. Under such conditions the term $G_{\mathrm{tx}}G_{\mathrm{rx}}$ of (2.3) becomes simply G^2 where G is the gain of the sole antenna. The RCS of the target is a measure of the target's size, *as seen by the radar*. Despite the unit of RCS being m^2 it depends on the backscattering properties of the target and is affected by target material and shape as well as the frequency of the radar signal. The pattern propagation factor is a measure of the gain, or attenuation, that occurs as a result of the route of the transmission. It is practical to contrast the received signal power with the system noise power to give an idea of the radar's maximum range and the minimum signal power that can be detected.

In order to be detectable the power of the backscattered pulse must have a greater power than the radar's noise floor. Thermal energy will cause there to be a low level output from the receiver chain even when there is no received signal present, this output is known as the noise signal. If the average noise power is symbolized by $\langle P_{\mathrm{n}} \rangle$ then the Signal to Noise Ratio (SNR) may be defined as:

$$\mathrm{SNR} = \frac{P_{\mathrm{rx}}}{\langle P_{\mathrm{n}} \rangle} = \frac{P_{\mathrm{tx}} G_{\mathrm{tx}} G_{\mathrm{rx}} \sigma \lambda^2 F^4}{\langle P_{\mathrm{n}} \rangle \, (4\pi)^3 R^4 L}. \tag{2.4}$$

The maximum range of the radar can be found by rearranging the expression for SNR for R:

$$R_{\max} = \left[\frac{P_{\mathrm{tx}} G_{\mathrm{tx}} G_{\mathrm{rx}} \sigma \lambda^2 F^4}{\langle P_{\mathrm{n}} \rangle \, \mathrm{SNR}_{\min} (4\pi)^3 L} \right]^{\frac{1}{4}} \tag{2.5}$$

where $\mathrm{SNR}_{\min}$ is the minimum SNR at which a signal may be detected. The maximum range given by (2.5) is the absolute maximum and depends purely on the power of the backscatter signal. The waveforms that are employed by radars to allow them to determine range impose further limits on the maximum range at which a target may reliably be detected. The above equations, (2.3) to (2.5), are examples of the radar equation that form the basis for the most fundamental radar theory and analysis.

2.2.1.1 Range ambiguity

The waveforms used to permit detection of range by a radar have a repetitive nature, e.g. a pulse is transmitted at a rate f_{PRF}, and the repetition interval imparts an upper limit on the radar's maximum range. In a pure Continuous Wave (CW) system the only limit on whether a target may be detected is the power of the backscatter signal and (2.5) holds true. Unfortunately it is not possible to determine target range with a pure CW signal since the lack of variation in transmission means that the time elapsed between the original transmission and the reception of the backscatter signal can not be determined. Introducing an Amplitude Modulation (AM) to pulse the signal, Fig. 2.3, permits the detection of range

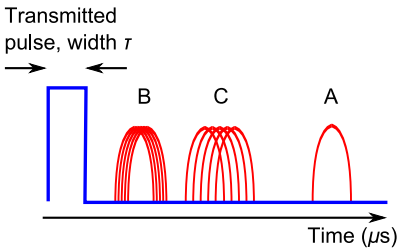


Figure 2.4: Output of a radar A-scope when multiple-time-around range ambiguity is in effect for three targets.

but limits the *maximum unambiguous range* of the radar to:

$$R_{\text{unamb}} = \frac{cT_{\text{PRI}}}{2} = \frac{c}{2f_{\text{PRF}}}. \tag{2.6}$$

If the round trip time to the target is greater than the PRI then the radar will have transmitted the next pulse before the backscattered signal has received. It will not be possible to tell which pulse the backscattered signal results from and it will be assumed to be a reflection of the last transmitted pulse resulting in an incorrect range measurement. The effect is known as multiple-time-around range ambiguity.

Varying the PRF changes the value of R_{unamb} from pulse to pulse causing the ambiguous range measurements to fluctuate while the unambiguous measurements remain constant. The apparent range of targets beyond R_{unamb} is a function of the unambiguous range so as the PRF, and hence R_{unamb}, fluctuates so do the ambiguous ranges. For targets at ranges less than the unambiguous range, that corresponds to the highest PRF, the range measurement is a function of the time delay only so their range measurement is constant. Fig. 2.4 shows the A-scope[1] output for a radar with a varying PRF illuminating three targets: the response from the unambiguous target A is now distinguishable from those targets, B and C, beyond R_{unamb} whose responses are spread in time.

The last factor affecting the range is the pattern propagation factor. It is a result of so called multi-path effects and may dramatically increase or decrease the radar's range.

2.2.2 Pattern propagation factor

The last variable of the radar range equation, (2.3), to be considered is the pattern propagation factor, F^4. It is possible that the route taken by the radar signal to and from the target is not the direct one of Fig. 2.2 but one that involves reflection from the ground plane or another large object. Considering only one reflecting object, the Earth or the ground plane, there are four possible paths between the radar antenna and the target:

[1] An A-scope is the most simple radar display. It consists of a cathode ray tube where the electron beam makes one horizontal sweep of the screen each PRI and the vertical height of the beam is controlled by the power of the received signal. As the beam sweeps across the screen it provides a range profile for the direction in which the antenna is pointing.

- The signal may travel directly to and from the target along a *direct path* and so cover a distance $2R_{\mathrm{d}}$, where R_{d} is the direct path range.
- Alternatively the signal may travel to and from the target along an *indirect path*, that includes the reflection from the Earth, covering a distance $2R_{\mathrm{i}}$, where R_{i} is the indirect path range..
- And lastly, the signal may travel to the target along the direct path and then return along the indirect path, or vice versa. For both of these situations the distance traversed is $R_{\mathrm{d}} + R_{\mathrm{i}}$.

The variation between the length of the indirect and direct paths results in a phase difference between the signals that traverse them; this can increase or decrease the amplitude of the signal when they combine according to the principle of superposition (Hecht, 1987). If the reflection coefficient for the indirect signal's reflected from the ground is assumed to be -1, indicating a perfect reflection with 180° phase shift, the value for F^4 is (Knott et al., 2004):

$$F^4 = 16 \sin^4 \left[\frac{k}{2} (R_{\mathrm{d}} + R_{\mathrm{i}}) \right]. \tag{2.7}$$

where $k = \frac{2\pi}{\lambda_{\mathrm{c}}}$ is the wavenumber. Since the $\sin^4(\cdot)$ expression may vary between 0 and 1 the value of F^4 increases the signal power between 0 and 16 times, a gain between $-\infty$ and 12dB. The extremes of this gain correspond to the case where the phase differences completely cancel the signal (180°out of phase) or where there is no phase difference resulting in a four fold signal amplitude increase, respectively.

The inclusion of the pattern propagation factor in (2.5), the calculation for maximum range, can have a dramatic effect. It is possible that the multipath completely cancels the backscatter signal, effectively hiding the target from the radar, or results in the detection of a target at four times the range possible when there is no multipath. Although such extreme variation may seem to make a mockery of the calculations to predict radar performance it must be remembered that a moving target will have a varying pattern propagation factor. As soon as the nearby target whose signal is completely cancelled moves the value of F^4 will change causing it to be seen, just as the distant target that is suddenly visible due to a four fold increase in $R_{\max}$ rapidly disappears from the radar scope.

2.3 Doppler Processing

The observed frequency of a wave of any kind, be it mechanical or electromagnetic, varies depending on the relative velocity between the observer and the source of the wave. This effect is known as the Doppler effect, and the change in frequency referred to as the Doppler shift, after the Austrian mathematician Christian Doppler who first explained the phenomenon. The effect is of particular significance in radar since both the radar and the target may be in motion and both may be regarded as the source of the wave through the transmission and backscattering process respectively. The most common approach to explaining the Doppler effect is to consider how moving the wave source can compress or stretch out the

wavelength that is analogous to raising or lowering the frequency. However, an alternative formulation is possible that starts by considering the phase of the received signal and views frequency as the time derivative of phase. This approach is used here since it is of value when understanding the micro-Doppler effect.

The phase of the backscattered signal depends on the range of the target; if the target is moving then this phase will vary with time. The change in phase between the transmission of the signal and the reception of the backscatter signal is a function of the distance to the target and back, $2R$, in wavelengths, λ_{c}, multiplied by the angular change of an oscillation over one wavelength, 2π radians:

$$\Delta\phi = 2\pi\frac{2R}{\lambda_{\mathrm{c}}} = \frac{4\pi R}{\lambda_{\mathrm{c}}}. \tag{2.8}$$

If the target is in motion relative to the radar then R will be replaced by a time dependent range term, $R(t)$ and it will be possible to take the time derivative of (2.8):

$$\begin{aligned}\frac{\mathrm{d}\Delta\phi}{\mathrm{d}t} &= \frac{4\pi\frac{\mathrm{d}}{\mathrm{d}t}R(t)}{\lambda_{\mathrm{c}}} \\ &= \frac{4\pi v_{\mathrm{r}}}{\lambda_{\mathrm{c}}}\end{aligned} \tag{2.9}$$

where the radial velocity of the target, v_{r} it's velocity along the radar Line Of Sight (LOS), has replaced $\frac{\mathrm{d}}{\mathrm{d}t}R(t)$ since velocity is the time derivative or range. The time derivative of phase is frequency so $\frac{\mathrm{d}}{\mathrm{d}t}\Delta\phi$ may be replaced by the angular Doppler shift ω_{d} leading to:

$$\omega_{\mathrm{d}} = -2\pi\frac{2v_{\mathrm{r}}}{\lambda_{\mathrm{c}}} \tag{2.10}$$

and the linear frequency version may be found through division of ω_{d} by 2π:

$$f_{\mathrm{d}} = -\frac{2v_{\mathrm{r}}}{\lambda_{\mathrm{c}}}. \tag{2.11}$$

In (2.10) and (2.11) a minus sign has been introduced. When a target is closing on the radar it introduces a positive Doppler shift, but has a *negative velocity* since range is measured *away* from the radar. By including a minus sign in the equations for Doppler shift a positive frequency shift is obtained for approaching targets. Equation (2.11) is the standard expression for the Doppler shift quoted in almost every radar text (Skolnik, 1980; Kingsley and Quegan, 1992; Stimpson, 1998). The measurement of f_d, and hence v_{r}, with a pulse radar places extra requirements on the transmitter and receiver. Unless *coherent* pulses are produced, the radar is unable to detect the Doppler shift.

Unless the Doppler frequency is substantially greater than the pulse bandwidth it is not possible to observe the change in frequency from a single pulse, and the Doppler shift must be determined over a

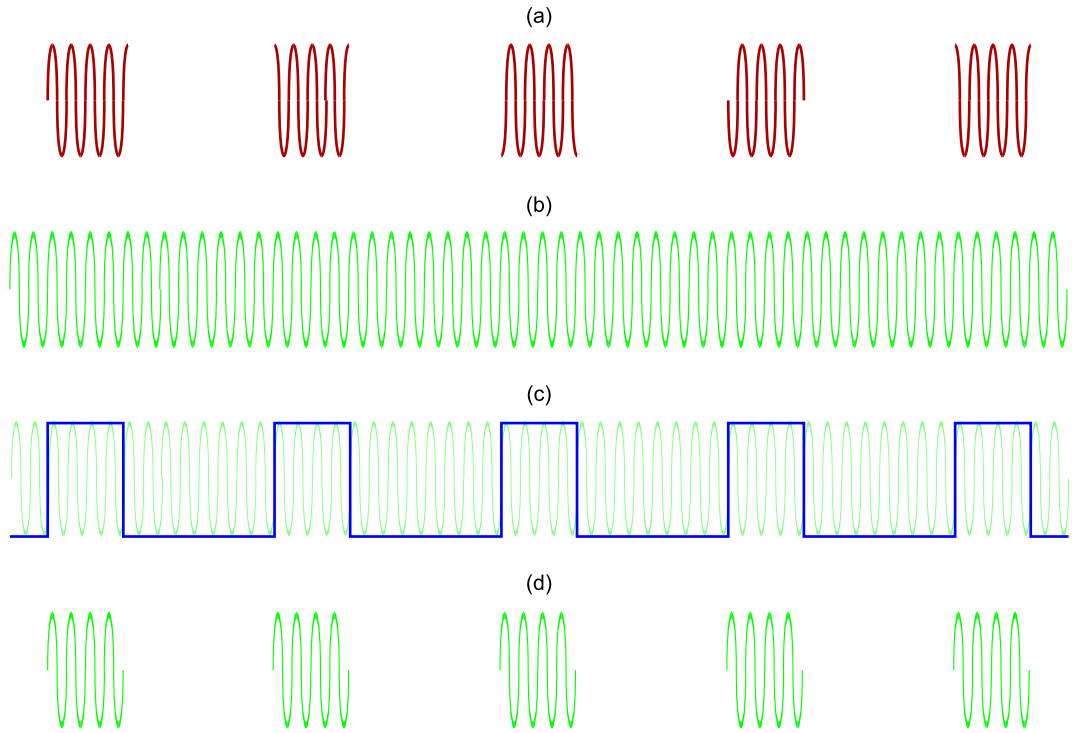


Figure 2.5: The process for generating coherent pulses. Part (a) is an example of incoherent pulses; (b) the output of the coherent oscillator; (c) the amplitude modulation to be imposed on the coho signal; and (d) the train of coherent pulses.

number of pulses. The phase of each received pulse differs from its transmitted phase according to the range of the target it was reflected from, as given by (2.8). If the target moves then the range is slightly different for successive pulses and it is possible to measure the rate of change of phase to calculate the Doppler frequency. However, for this monitoring of phase change to work *the transmission phase of each pulse must be known*. If the transmission phases are unknown then it is not possible to determine how much the phase has changed. Modern system achieve this knowledge by "cutting" the pulses from a continuous signal so that each has the same initial phase, a property known as coherency.

In order that a pulse train be regarded as coherent, it must adhere to the following definition:

> *The first wavefront of each pulse must be separated from the last wavefront of the same polarity in the preceding pulse by a whole number of wavelengths.* (Stimpson, 1998)

The process for generating a coherent train of pulses is shown in Fig. 2.5. For comparison part (a) of the figure shows five incoherent pulses in which the initial phase of each pulse is different. Part (b) shows the continuous signal that is generated by the coho and part (c) shows the AM signal or, pulse envelope, that is used to control the power amplifier in Fig. 2.1 superimposed on the coho signal. The last part of the figure, part (d), shows the coherent pulse train generated by this method: the initial phase of each of these pulses is the same. The steps of Fig. 2.5 differ slightly from the description of section 2.1 because there is no up-conversion performed. This has been omitted for clarity since, by design, the up-

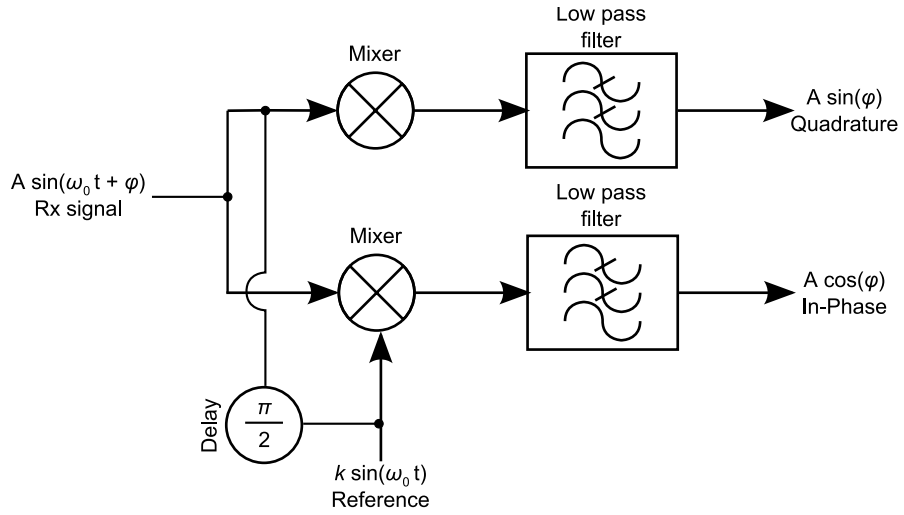


Figure 2.6: The quadrature detector.

and down-conversions from mixing with the stalo signal do not effect the phase of the reference signal. Radars that emit coherent pulses to measure the Doppler shift are known as pulse-Doppler radars.

The pulse-Doppler radar measures the distribution of energy across the spectrum of the received signal to measure the radial velocity of targets and to separate targets at the same range. Equation (2.11) details how the velocity of the target relative to the radar shifts the frequency of the backscatter signal. By measuring the frequency shift it is a simple calculation to work out the velocity of the target along the radar LOS. Furthermore, if two targets at the same range are travelling with different velocities then their individual backscatter signals will have different Doppler shifts permitting the separation of the targets. In order to achieve the measurement of target velocity the radar requires a bank of bandpass filters, to detect which frequencies are present in the signal, and that the sign of the phase of the received pulses is known. The later of these requirements is met through the use of the quadrature detector.

The quadrature detector is two synchronous detectors working in tandem that outputs the phase of the input signal and the phase of the input signal delayed by 90°; this permits detection of the sign of the phase. The structure of the detector is shown in Fig. 2.6. The received signal is described by the expression $A\sin(\omega_0 t + \phi)$ where $\omega_0 = 2\pi f_0$ is the angular reference frequency of the radar and ϕ the phase shift on a pulse resulting from target motion. They synchronous detector removes the reference signal from the received pulse leaving only the phase change. The mixer stage represents a multiplication of the signals leading to:

$$\begin{aligned} V_{\text{output}} &= A\sin(\omega_0 t + \phi)\,.k\sin(\omega_0 t) \\ &= Ak\sin(\omega_0 t)\left[\sin\phi\cos(\omega_0 t) + \cos(\phi)\sin(\omega_0 t)\right] \\ &= \frac{kA}{2}\left[\sin(\phi)\sin(2\omega_0 t) + \cos(2\omega_0 t) + \cos(\phi)\right]. \end{aligned}$$

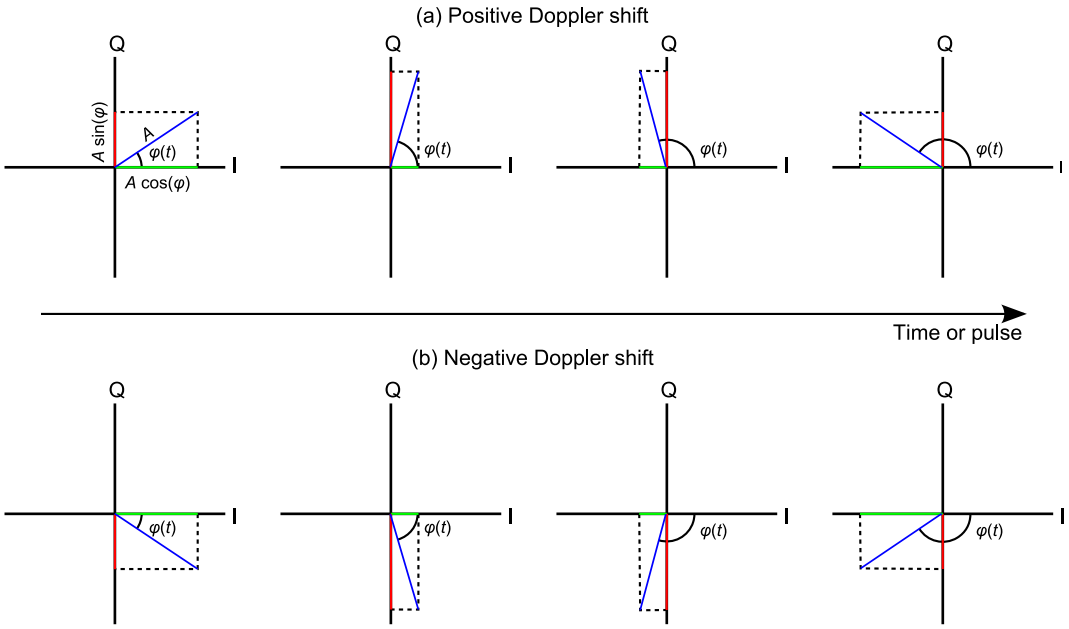


Figure 2.7: Phasors showing the effect of (a) positive and (b) negative Doppler shifts on the progression of the phase with time.

The low-pass filter removes the expressions representing frequencies of $2\omega_0$, and through arranging for k to equal 2 the output of the synchronous detector is:

$$V_{\text{output}} = A\cos(\phi)\,, \tag{2.12}$$

which is the in-phase, or I-channel, output of the quadrature detector. If the phase of the reference signal fed into the detector is delayed by 90° or $\frac{\pi}{2}$ radians then the output is:

$$\begin{aligned} V_{\text{output}} &= A\cos\left(\phi - \tfrac{\pi}{2}\right) \\ &= A\sin(\phi)\,, \end{aligned} \tag{2.13}$$

which is the quadrature, or Q-channel, output. These two outputs can be thought of as the axes of a graph allowing the importance of having the two channels to be explored graphically.

Fig. 2.7 shows how the phase of the received signal varies with time (or successive pulses) through the phasor representation. The reference for the phasor is an oscillation of frequency f_0 so the received signal is a line vector of length A at angle ϕ to the I-axis. In part (a) of the figure the development of the phase is seen when there is a positive Doppler shift. As successive pulses are received the phase difference at each pulse increases and the phasor rotates in an anti-clockwise direction. The projection of the phasor onto the I- and Q-axis show the output of the in-phase and quadrature channels of the quadrature detector. Part (b) of Fig. 2.7 shows the phasor progression when the Doppler shift is negative. This time the rotation is in a clockwise direction with time, but the projection onto the I-axis is the

same as in the positive Doppler case. Unless the Q-channel information is available it is only possible to determine that the target is in motion and the speed with which it is moving, and not whether it is travelling towards or away from the radar. The effect is known as *fold over*: unless both I- and Q-channel information is available negative frequencies are "folded" about 0 Hz and appear as positive frequencies. For convenience the in-phase and quadrature channels can be considered as the real and imaginary parts of a complex number. This leads to the complex notation for the baseband signal, $v(t)$:

$$\begin{aligned} v(t) &= A\cos(2\pi f_{\mathrm{Dop}}t) + iA\sin(2\pi f_{\mathrm{Dop}}t) \\ &= A\cos(2\pi f_{\mathrm{Dop}}t) + iA\sin(2\pi f_{\mathrm{Dop}}t) \\ &= Ae^{i2\pi f_{\mathrm{Dop}}t}, \end{aligned} \tag{2.14}$$

where f_{Dop} is the Doppler frequency. The complex signal notation is particularly convenient for explaining how filters are able to deconstruct the signal into its varying frequency components.

Any periodic signal can be deconstructed in to a series of pure tones, its Fourier series, permitting analysis of the frequencies present in the signal (Proakis and Manolakis, 1996). If the signal is $v(t)$ then the Fourier series is:

$$v(t) = \sum_{-\infty}^{n=\infty} c_n e^{i2\pi f_n t} \tag{2.15}$$

where the constant c_n is defined as:

$$c_n = \frac{1}{2\pi(\Delta t)} \int_{t_1}^{t_2} v(t) e^{-i2\pi f_n t} \mathrm{d}t, \tag{2.16}$$

t is time, f_n is the n^{th} frequency component, t_1 and t_2 the start and end time of the signal, and Δt the duration of the signal, $t_2 - t_1$. The constant c_n is a complex number whose amplitude and argument give the amplitude and phase, respectively, of the signal corresponding to the n^{th} frequency component of $v(t)$. A bandpass filter allows through a range of frequencies surrounding its centre frequency making it possible to estimate the amplitude of the constant c_n for the centre frequency of the filter. A filter bank is a set of contiguous filters that allow the estimation of the amplitude of c_n for a range of frequencies. Before considering the performance of the filter bank, the performance of a single filter should be considered.

The frequency spectrum of the bandpass filter defines the range of frequencies it will pass, *the bandwidth*, and the attenuation that will be applied to those frequencies. Ideally the filter would only pass a single frequency, but this could only happen if the signal were applied to the filter for an infinite amount of time i.e. Δt in (2.15) and (2.16) is infinite. The extremely short width of a single radar pulse, τ, prevents the measurement of the frequency components of the signal from just one pulse. Instead a pulse train must be used. Fig. 2.8 compares the spectra of a bandpass filter for one pulse and a train of pulses. In part (a) of the figure the video (envelope) of a single pulse is shown, it is much longer than a

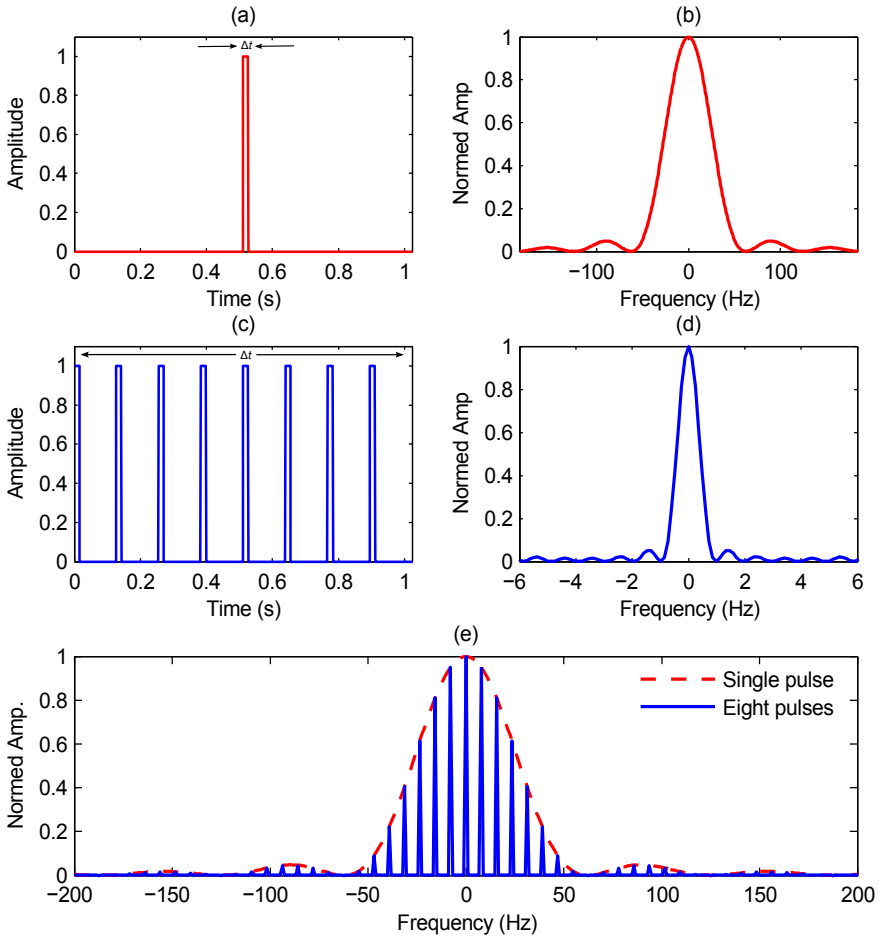


Figure 2.8: The frequency spectrum of a bandpass filter: parts (a) and (b) show the video of a single pulse and its frequency spectrum while parts (c) and (d) show the video and spectrum of a train of pulses. Part (e) shows the full spectrum of the pulse train including the ambiguous frequency pass regions and how their amplitude is modulated by the spectrum for a single pulse.

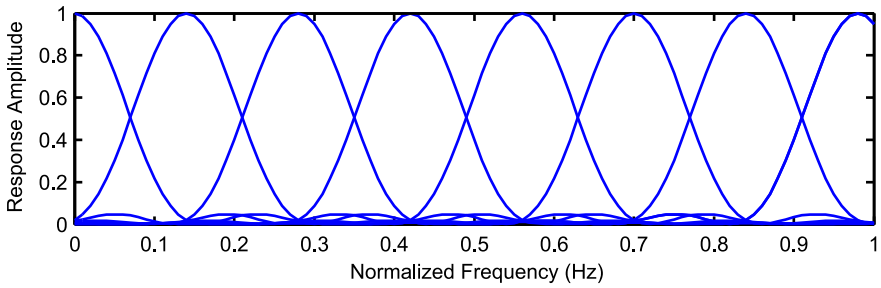


Figure 2.9: A bank of eight bandpass filters covering a frequency range.

typical radar pulse being 16ms long but still serves to illustrate the point. The spectrum of a bandpass filter to this signal is shown in part (b) of the figure and is seen to be the square of the sinc function.

$$\mathrm{sinc}(f) = \begin{cases} 1 & \text{if } f = 0 \\ \frac{\sin(\pi f)}{\pi f} & \text{if } f \neq 0. \end{cases} \tag{2.17}$$

In this instance the filter is centred on 0 Hz, but it could be centred on any frequency depending on the filter being used. The width of the main lobe, and all subsequent lobes of $\mathrm{sinc}^2(f)$, relates to the duration of the pulse and is $\frac{2}{\Delta t} = 2B$ where B is the pulse bandwidth equal to $\frac{1}{\tau}$. In the example the distance between nulls in the filter spectrum is 125 Hz. In part (c) of the figure a train of eight pulses is shown. Δt is now the duration of the pulse train, 1.024 s in the example, and the null-to-null width of the filter is now ≈ 2 Hz. This is a substantial improvement on the single pulse case, but the zoomed in frequency axis of part (d) does not provide the full picture. In part (e) much more of the frequency spectrum of the train of pulses can be seen as the solid blue line. It has a repeating structure with a large peak in the spectrum every ≈ 8Hz, which is the repetition frequency of the train of pulses from part (c) of the figure. The amplitude of these repeating peaks in the spectrum is modulated by the spectrum of the single pulse, shown by the red dashed line. By suitably positioning in frequency a bank of such bandpass filters the significance of each f_n from (2.15) can be estimated.

By overlapping a large number of bandpass filters the frequency range of interest can be completely covered. Fig. 2.9 shows such a filter bank in which eight bandpass filters have been used to cover a frequency range. The filters are overlapped such that the spectra cross when the amplitude has fallen to 0.5, a drop of 3 decibels (3 dB). If the received signal contains frequency components that are in between the peaks of the filter's spectrum then the filters either side of the frequency will have an output. By interpolating between the two filters it is possible to work out the precise frequency of the signal component. It is common for the 3 dB bandwidth to be used as the description for a filter and, conveniently, the 3 dB bandwidth of a bandpass filter is equal to $\approx \frac{1}{\Delta t}$. The relationship between the filter bandwidth and duration of the signal applied couples the effectiveness of the filter bank to the signal time Δt. If the signal is applied for too short a time then the filters broaden, the 3 dB bandwidth will increase, and

it will become difficult to determine the precise frequencies of the signal components since they may be passed by two or more filters evenly. Conversely, if the signal is applied for too long then the filters will become overly fine, gaps will appear between them and frequency components of the received signal will be "lost" in the gaps. Deciding on the required frequency resolution and duration of measurement period is critical when designing a filter bank. Achieving suitable resolution can require a large number of expensive filters leading to the main drawback of analogue Doppler techniques.

The Discrete Fourier Transform (DFT) provides a mechanism to investigate the frequency components of a digitized signal without using a filter bank. Instead, the signal is digitized using an Analogue to Digital Converter (ADC) resulting a discrete, rather than continuous, signal. For a discrete signal $x[t]$ the DFT is represented by $X[f]$ and the two are related by:

$$X[f] = \sum_{t=0}^{N-1} x[t] e^{-i\frac{2\pi f t}{N}} \quad \text{where} \quad \{f \in 0 \cup \mathbb{N} : f < N\} \tag{2.18}$$

and

$$x[t] = \frac{1}{N} \sum_{f=0}^{N-1} X[f] e^{i\frac{2\pi t f}{N}} \quad \text{where} \quad \{t \in 0 \cup \mathbb{N} : t < N\} \tag{2.19}$$

where f is frequency and t time; and N the number of samples of the signal. The direct calculation of the DFT on a computer requires of the order N^2 computations, making it computationally intensive, but the famous Fast Fourier Transform (FFT) algorithm exploits symmetry in the calculation to reduce the computational overhead to of the order $N \log N$ computations allowing quick calculation of the frequency spectrum (Gauss, 1866; Cooley and Tukey, 1965). Before contrasting the DFT with the analogue filter bank its properties must be understood.

The sampling frequency, f_{sample}; number of samples, N; nature of samples, real or complex; and the signal being sampled all affect the spectrum produced by the DFT. The signal of interest will be comprised of a number of frequency components, and the highest of these that may be reliably sampled is given by the Nyquist frequency. For any discrete sampled waveform the highest frequency component must be no larger than $f_{\text{sample}}/2$ if real samples are used. If complex samples are used, the highest frequency component of the waveform may be f_{sample}. Frequency components above this Nyquist frequency will be aliased and appear in the range 0 Hz to f_{Nyq} Hz at a frequency that could generate the same set of samples. The number of samples taken from the signal governs the number of frequency bins produced by the DFT, and their width is $1/\Delta t$ where $\Delta t = N/f_{\text{sample}}$. The frequency range spanned by the DFT output is described by the interval $[0, f_{\text{sample}})$. If the real valued samples are used then there is only valid information in the interval $[0, f_{\text{sample}}/2)$; the remaining interval, $[f_{\text{sample}}/2, f_{\text{sample}})$, contains a mirror of the spectrum from the first interval. If complex samples are used then the full frequency span of the DFT contains useful information. Should the DFT be used to investigate a signal

that contains a negative frequency, such as when the Doppler shift from a radar target is investigated, the negative frequencies appear as if added to the sampling frequencies i.e. for a shift of $-f_{\mathrm{Doppler}}$ the spectral component produced by the DFT will be at $f_{\mathrm{sample}} - f_{\mathrm{Doppler}}$ it is common therefore to swap the position of the intervals $[0, f_{\mathrm{sample}}/2)$ and $[f_{\mathrm{sample}}/2, f_{\mathrm{sample}})$ output by the DFT and consider the output as covering the interval $[-f_{\mathrm{sample}}/2, f_{\mathrm{sample}}/2)$. It is the final property of the DFT output, the effect of the signal window, that makes it truly comparable with a bank of bandpass filters.

The set of N samples to which the DFT is applied provide a finite duration representation of the signal causing a sinc function to appear in the output spectrum just like a bandpass filter. Theoretically the application of a Fourier transform to a pure tone should result in delta-function spike in the frequency spectrum, but this assumes the signal to be of infinite duration. In practice the signal will always be of finite duration and can be considered as being comprised of an infinite duration pure tone that is amplitude modulated by a rectangular function covering the interval $[t_1, t_2]$. The rectangle function is defined as (Hrynkiewicz, 1998):

$$\mathrm{rect}(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if } t < 0 \text{ or } t > 1 \end{cases} \tag{2.20}$$

for a rectangle of duration 1; if a rectangle of some other duration, τ, is required $\mathrm{rect}(\frac{t}{\tau})$ can be used. So the signal to which the DFT is applied can be considered as:

$$v(t) = \mathrm{rect}\left(\frac{t}{t_2 - t_1}\right).\sin(2\pi f t). \tag{2.21}$$

Since a multiplication in the time domain is equivalent to a convolution in the frequency domain the resulting spectrum after sampling and application of the DFT to $v(t)$ will be a convolution of the spectrum for a pure tone with frequency f, a delta spike at f, and the spectrum of the rectangle function of duration τ that is a sinc function with 3 dB bandwidth $\approx 1/\tau$. The final spectrum is then a sinc function centred at f since the convolution of any function with a delta function is the original function centred at the location of the delta function. By considering each of the N frequency bins created as output of the DFT to be the position at which the delta function would occur for a particular f_n the resulting spectrum is identical to that obtained using a bank of N bandpass filters to which the signal was applied for a duration $t_2 - t_1$. For a signal of duration $\approx$ 1s an 8 sample DFT will be comparable with the filter bank shown in Fig. 2.9. The DFT approach has advantages over the physical filter approach, however, since it does not require any microwave filters. The implementation of the FFT algorithm on a digital processor means that the only limit on the number of filter banks available is a function of processor speed making one bank of filters, implemented as one DFT, per range bin practical.

The advances in computer processing power in recent years has permitted the wide spread devel-

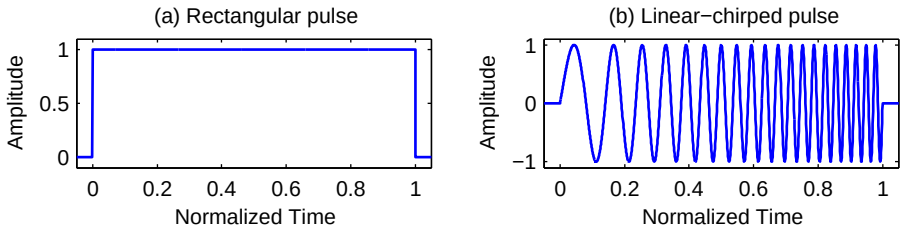


Figure 2.10: The shape of the video signal for (a) rectangular pulses and (b) chirped pulses.

opment of cheap pulse-Doppler radars as the required DFTs can run quickly. In fact, digital signal processing techniques are becoming more and more common in radar and many RF stages can now be replaced with digital methods. This trend permits radars to be much more flexible since it is relatively easy to alter the digital processing chain during operation compared with the RF hardware.

The final part of the radar theory to consider is the waveform. This critical part of the radar design controls the range resolution and the Doppler response of the system.

2.4 Radar Waveforms

The range resolution of a pulse radar was described in (2.2) as a function of the PW. The definition can also be given in terms of the pulse bandwidth:

$$\Delta R = \frac{c\tau}{2} = \frac{c}{2B} \tag{2.22}$$

where c is the speed of propagation, τ the PW and B the pulse bandwidth that is equal to $1/\tau$ for a rectangular pulse. It is clear from the equation that shorter pulses give better resolutions i.e. a smaller ΔR. Unfortunately it is not possible to keep decreasing τ. The radar's transmission power is averaged over the transmitted pulse duration; the shorter the pulse the greater the instantaneous power that is required. Fortunately ΔR can also be considered as a function of the bandwidth, B, and this may be increased without reducing the PW, by varying the frequency during transmission.

A linear increase in frequency during transmission can improve the radar range resolution provided a suitable filter is applied to the received signal. The entire receiver channel can be considered as a filter; the frequency, phase and amplitude response of this filter will affect the output signal. The first consideration is to maximize the SNR since this maximizes the chance of detecting a target. A filter whose properties exactly match the transmitted pulse achieves the maximum SNR in its output. Such a filter is referred to as a *matched filter* in radar literature since it matches the transmission. The output of the matched filter is a cross correlation of the transmitted pulse with the received signal. Here the output of the matched filter for two types of pulse will be discussed, the regular rectangular pulse and the linear chirped pulse, both of which are shown in Fig. 2.10. The two examples are for the bipolar video signal

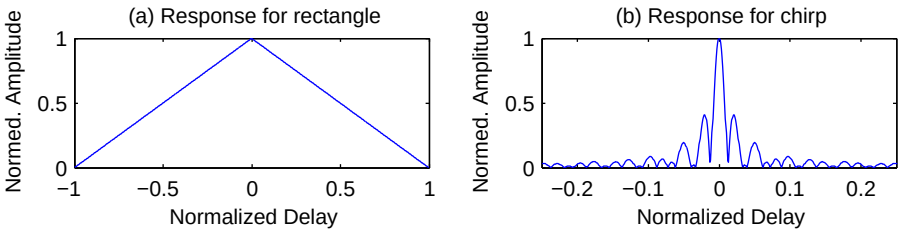


Figure 2.11: The response of a matched filter for (a) a rectangular pulse and (b) a chirped pulse.

since the matched filter will be applied after the reference oscillation has been removed i.e. the matched filter is applied after the quadrature detector in Fig. 2.1. The envelope of the unchirped pulse, part (a) in the figure, is a standard rectangle. The chirped pulse, part (b) in the figure, is a linearly increasing oscillation. The application of a matched filter to these signals results in two different and distinctive responses.

Fig. 2.11 shows the matched filter response for the two pulses from Fig. 2.10. (It is important to remember that the matched filter is different in each case since it is matched to pulse.) In this situation the filter response is effectively the autocorrelation function. Part (a) of Fig. 2.11 show the triangle function that is obtained when a rectangle is cross correlated with itself. The width of the triangle at the half-amplitude, 3 dB, point is ≈ 1 so if two pulses were overlapped at this point their peaks would be separated by ≈ 0.5 in agreement with (2.22). When the pulse has a linear frequency chirp the matched filter response is as shown in part (b) of Fig. 2.11. The filter response is now the distinct $\mathrm{sinc}^2(\cdot)$ function and the width of the mainlobe at the 3 dB amplitude point is a small fraction of the regular rectangle. It is clear that the targets can be much closer together and still distinguished from one and other when using this waveform. In these examples it has been assumed there is no Doppler shift of the received signal; if there were a shift then the response of the filter would be different. To analyse the filter response in terms of both time delay and Doppler shift the *radar ambiguity function* is required.

The radar ambiguity function (Levanon and Mozenson, 2004) permits the investigation of the frequency and range response of a matched filter simultaneously. It is defined as:

$$\chi(\tau,\nu) = \int_{-\infty}^{\infty} u(t)u^*(t-\tau)e^{i2\pi\nu t}\mathrm{d}t \tag{2.23}$$

where τ and ν are time delay and frequency shift, respectively, t is time, $u(t)$ the signal and $u^*(t)$ the complex conjugate of the signal. Since the function will be complex it is common to plot the absolute value $|\chi(\tau,\nu)|$. Fig. 2.12 is the ambiguity diagram, the plot of $|\chi(\tau,\nu)|$, for a train of rectangular pulses. The x-axis shows the delay and the y-axis the frequency shift; the units of these axes have been normalized by the duration of the pulse and the number of samples used in the simulation. The plot has rotational symmetry about the point (0,0) but the negative frequencies have been omitted to allow the

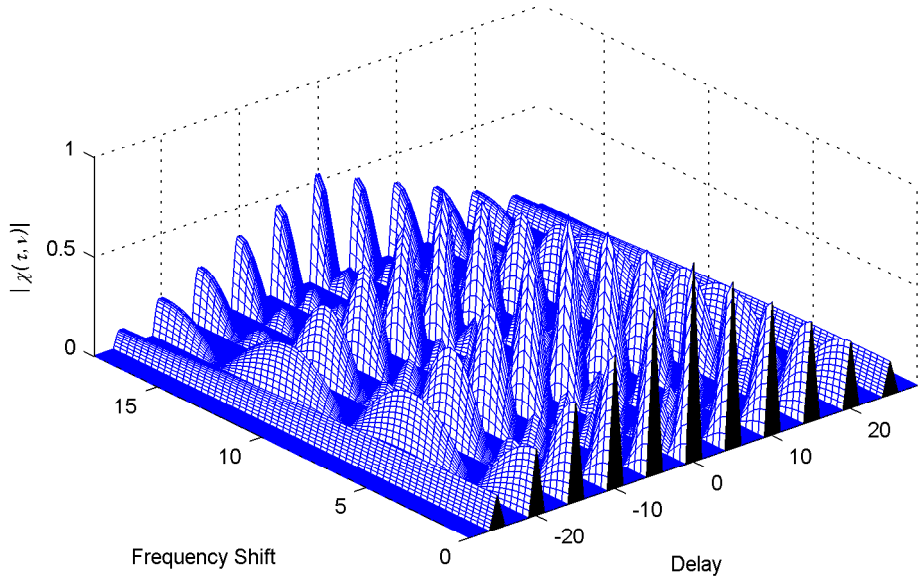


Figure 2.12: The ambiguity diagram for a train of six rectangular pulses.

zero-Doppler cut to be seen, this is the autocorrelation that occurs when there is no Doppler shift. The ambiguity diagrams show that although the target maybe well located in position and velocity (measured as frequency) in the real world their response in the radar is distributed leading to the possibility of inaccurate detection or masking of one target by another. Clearly the waveform used should be matched to the radar's intended function.

The most common use for waveform design is to improve the radar's range resolution and there are many approaches that may be taken to achieve this. In the above descriptions the focus has been on using a Frequency Modulation (FM) chirp to achieve improved resolution, but it is just as valid to use pulse coding. By suddenly changing the phase of the transmitted signal binary digits may be encoded onto the waveform and these too improve range resolution. Common encoding are Barker and Costas codes to improve range resolution while Deng codes are employed to create orthogonal encodings for use when several pulses must be transmitted simultaneously. The radar waveform may also encompass multiple pulses. In a stepped frequency waveform the transmit frequency is increased over a number of pulses. These pulses are then processed together and the bandwidth is considered as being the change in frequency across the group of pulses. Through such methods extremely high range resolutions may be achieved and theses techniques are often used in High Range Resolution Profile (HRRP) techniques, one of the main alternatives to Micro-Doppler Signature (μ-DS) for target recognition.

Now that a suitable base of radar theory has been developed it is possible to move on and discuss the current state of radar Automatic Target Recognition (ATR), including techniques such as HRRP. It must be remembered, however, that this chapter on radar basic theory has been very top-level and is in

no way a complete description of all aspects of radar. Despite this the reader should now be equipped with sufficient radar knowledge to make sense of the research described in this thesis. Where it is needed further theory will be discussed *in situ*.

2.5 Literature Survey

2.5.1 Bistatic And Multistatic Radar

Most current radar systems are monostatic in nature: the transmitter and receiver are located at the same geographic location and often share the same antenna (Swords, 1986; Skolnik, 1990). Radars of this type have allowed great advances in remote sensing capability. This includes the detection of targets in adverse weather conditions and beyond the horizon as well as the imaging of distant planets (Stimpson, 1998). However, there is another class of radar which can provide additional functionality.

Multistatic radars, of which the bistatic radar is a special case, are systems comprised of multiple transmitters and receivers that are geographically distributed but operate as one radar (Willis, 2005; Cherniakov, 2007; Chernyak, 1998). While these systems are more complex to develop they offer potential gains over their monostatic counterparts. The most often cited gains for multistatic radars are: their potential robustness to anti-radiation missiles—receive only nodes can not be detected by the missile; their immunity to conventional jamming since a single jammer will be unable to jam all the channels of the radar; and their potential for detecting stealth aircraft since these are only designed to be stealthy in the monostatic case. Further advantages of multistatic radars depend on the way their data is processed. With suitable algorithms it is possible for targets to be located with sub range-resolution cell accuracy (Doughty et al., 2007; Sammartino et al., 2008). Despite the increased complexity of these systems they may still be developed extremely cheaply using commercially available hardware (Derham et al., 2007, 2006; Doughty, 2008).

Within the context of radar ATR multistatic radars are of relevance as they permit multiple perspectives of the same target to be viewed simultaneously. This capability is utilised heavily in the investigation of Chapter 8.

2.5.2 Radar Automatic Target Recognition

2.5.2.1 Target Classification

The notion of wishing to recognize an object from the output of a sensor is in no way unique to radar. There is long established field of computer science dedicated to such work known broadly as *classification*. There are many text books describing this field. Of particular use in this research was *Pattern Classification* by Duda, Hart and Stork (Duda et al., 2005) that gives an excellent introduction to the most popular classification approaches. However, despite the broad nature of this work Dynamic Time Warping (DTW), a technique permitting the classification of time domain sequences, is not covered.

DTW is a classification method originally used in speech processing that shows potential for classifying radar signals. The technique was specifically designed to process time series data (Sakoe and Chiba, 1978) that is comparable to the output of a radar from a single range bin over a series of pulses. In fact, this treatment of the radar output is comparable to conventional pulse-Doppler processing where each pulse is considered as a sample of a time varying signal (Stimpson, 1998; Kingsley and Quegan, 1992; Skolnik, 1990, 1980). The computational overhead associated with DTW is known to be large (Myers et al., 1980) and this caused it to fall from prominence in speech recognition. However, a series of recent advances in both optimization of the algorithm and its software implementation have resulted in the technique re-emerging in recent years (Keogh and Pazzani, 2001b,a; Keogh and Ratanamahatana, 2002). As a result of these developments it was expected that DTW would be a valuable method for radar ATR.

Two further classification approaches were considered as part of this investigation: the k-Nearest Neighbour (k-NN) and the naïve Bayesian classifiers. Both of these are well documented in the general literature, (Duda et al., 2005) for example, and are regarded as well established approaches to classification. The high success rate of the naïve Bayesian approach is often regarded as surprising due to the highly simplified, or naïve, assumptions taken in formulating the classifier. However, detailed analysis of why these naïve assumptions result in such a power classifier has been performed and it has been demonstrated that the classifier may be used without concern about the assumptions (Hand and Yu, 2001; Zhang, 2004). (Further support for the naïve Bayesian classifier is given by Prof. David Hand of Imperial College London—the author of(Hand and Yu, 2001)—in the e-mail reproduced in Appendix A.)

2.5.2.2 Radar Target Recognition

The notion that a radar should be able to identify its targets as well as detect them is as old as radar itself and tentative efforts were made to achieve target recognition during the Second World War (Swords, 1986). These initial attempts were limited by the degree of signal processing required and it was decided to use transponders, rather than radar ATR, to distinguish friendly and hostile targets. However, active research into achieving automatic recognition continued and by the time the second edition of *Introduction To Radar Systems* was published in 1980 Skolnik was able to outline ten different phenomena upon which radar ATR might be based.

The earliest attempts at performing radar ATR were based on identifying the fluctuations in target RCS. The famous diagram of the variation of RCS with aspect angle for a B-26 bomber (Skolnik, 1980, 1990) demonstrates that an estimate of the RCS fluctuation would provide a classifiable target signature (Turhan-Sayan, 2005; Jofre, 1993; Silverstein et al., 1991). Despite the longevity of this approach its significance has been reduced in recent years by the rise in HRRP, Inverse Synthetic Aperture Radar (ISAR) and Synthetic Aperture Radar (SAR) based techniques.

HRRP classification has achieved impressive results and been the subject of much investigation (Zyweck and Bogner, 1995; Rihaczedk and Hershkowitz, 2000). However, the technique is known to be very sensitive to the aspect angle of the target (Liao et al., 2000) since its feature vectors rely on the relative position of scattering centres along the radar LOS. As Zyweck notes, if multiple HRRPs are available with suitable angular separation between them then an ISAR image of the target maybe formed, and this too can be used as the input to an ATR system (Kim et al., 2005). Research into ISAR based radar ATR often relies on turntable ISAR images where the data is generated by placing the target on turntable and rotating it through 360° in very small angular increments to generate a high quality image (Liao et al., 2001). While these images may lead to very sophisticated feature extraction and classification approaches, e.g. syntactic based classifiers (Vespe et al., 2006b), it must be remembered at all times that turntable ISAR images are ideal cases and the images in operational situations will be of reduced quality. An alternative to ISAR images are regular SAR images of a target.

In SAR based ATR small sub-images, known as "clips", that contain a specific target are cut from the main SAR image and classified (Schumacher and Schiller, 2005). Since SAR images are very similar to images obtained by optical methods classification techniques from disciplines such as machine vision can be directly utilized to form the ATR system (Gamba et al., 2000). The correlation between SAR and optical images is not perfect, however. Additional research has demonstrated that the characteristic target shadows in SAR images are as valuable a signature as the target image (Jahangir et al., 2007). Although the results that SAR based radar ATR achieve are very impressive it has shortfalls in certain areas. The SAR processing routines are complicated and have substantial computation overheads to form an image. The radars required to produce SAR images generally require high range resolution making them expensive to produce. Lastly, a conventional SAR image may cover a large geographical area leading to further potential drawbacks. It can take some time to survey the area covered and, assuming the radar is mounted on an aircraft, obtaining a new image of the same region requires the aircraft to fly past again limiting the update rate of situation on the ground. Furthermore, since the SAR image covers a large area it is quite possible that many targets are included within it, so in addition to the challenge of classifying targets the ATR system must incorporate a method to identify "clips" that contain candidate targets.

The radar ATR techniques discussed so far all rely on classifying RF radiation phenomena that relate to the physical structure of the target. Be it through the direct recognition of the RCS that corresponds to the physical measure of the target, as seen by the radar, or through the classification of an image of the target in techniques such as HRRP, ISAR and SAR image classification. However, there is another category of radar ATR solution that relies on classifying a target's Doppler response. These methods use the Doppler responses of sub-sections of the main target as a distinguishing target signature. This μ-DS radar ATR has been applied most widely in the recognition of Jet Engine Modulation (JEM). However,

recent advances in hardware and signal processing mean that now even the μ-DS of a walking person target is detectable and may be used as the basis for target recognition.

2.5.2.3 Micro-Doppler Signature ATR

The JEM phenomena only applies to air targets and is the result of the radar signal being reflected by the blades of the jet engine turbine as well as the fuselage. Under such conditions the frequency response of the target becomes a series of discrete spectral lines (Schneider, 1987; Martin and Mulgrew, 1990, 1992; Tong et al., 1998). The overall bandwidth of the backscattered signal, the spacing of the spectral lines and the relationship between the power of the upper and lower regions of the frequency response about the Main Doppler Line (MDL) all relate to the properties of the blades in the jet engines. Since it is possible to model the frequency distribution of the backscatter signal, prototype power spectra for candidate targets' responses can be developed and simple cross-correlation can be used to see which prototype matches the signal received by the radar. More recent attempts at performing recognition of the JEM for a target have focused on using Artificial Neural Networks (ANNs) that learn directly from pre-recorded data and so forgo the need to have accurate equations to describe the expected spectral response of the backscatter signal (Cuomo et al., 1994; Pellegrini et al., 1995; Piazza, 1999). It is also possible to detect and identify helicopters based on the frequency modulation induced by their rotor blades using similar techniques (Martinez Madrid et al., 1992; Misiurewicz et al., 1998; Johnsen et al., 2003). However, it is not only air targets that exhibit a characteristic micro-Doppler response. As radar equipment became more sensitive it became possible to detect frequency responses from target components such as wheels, caterpillar tracks and even swinging limbs. The theory of the micro-Doppler effect was developed to describe these responses and it too provides a unique signature by which a target may be identified.

Before considering the attempts that have been made to classify the μ-DS it is worth considering the development of the theory explaining the phenomenon. As noted above, initial efforts focused on the signatures associated with jet engine blades and helicopter rotors. From the year 2000 to the present day Victor Chen of the Naval Research Laboratory, Washington DC, has produced a series of publications detailing mathematical approaches to explaining the μ-DS. Initially these focused on describing the signatures associated with the basic types of micro-motion a target component may exhibit (Chen, 2000; Chen and Ling, 2002; Chen et al., 2003, 2006). While these works provide an excellent description of the underlying theory, they do not explain the μ-DS for any specific target. Since 2006 there have been a series of papers published that explore the μ-DS for a single class of target: personnel. Preliminary work developed a model that could be used to assist in the classification of personnel observed by through-wall radar (van Dorp and Groen, 2003). However, at the 2008 IEEE Radar Conference several papers were presented that focused on simulating and understanding the μ-DS generated by personnel (Ghaleb et al., 2008; Ram and Ling, 2008; Chen, 2008; Smith et al., 2008b). A common theme in these papers is their

interest in the different types of motion a person may exhibit—an interest motivated by contemporary concerns about terrorism and the changing face of warfare, in conflicts such as that in Iraq, requiring that human intent be determined. Before turning to the way the μ-DS may be classified it is worth noting that micro-Doppler is most commonly analysed using the Short Time Fourier Transform (STFT). It is necessary to use a time-frequency transform as the frequency modulation induced by the micro-Doppler effect varies with time. Although there are time-frequency transforms that provide better spectral resolution than the STFT all of these introduce so called "cross-terms" that are virtually impossible to distinguish from the low power μ-DS components of the power spectra (Marple, 2001, 2003).

μ-DS based radar ATR research has already resulted in its incorporation into a commercial radar (Stove and Sykes, 2003; Stove, 2006). Another approach that could be implemented on a real system (Bilik et al., 2006) has been published and a series of papers have been produced that have furthered the understanding of how practical classifiers may be developed in the future (Jahangir et al., 2003b,a; Martinez Madrid et al., 1992; Jiajin and Chao, 2005; Jiajin, 2005). The commercial system (Stove, 2006) uses linear discrimination of the μ-DS to decide between three ground target classes alongside platform specific functionality to determine a further three classes. Theoretical research shows the μ-DS to be time varying (Chen et al., 2006) although this is not taken into account in the commercial system as the radar dwell time is regarded short enough that the μ-DS may be considered constant over a single target illumination. Exploitation of the time variance has been attempted through the use of a Hidden Markov Model (HMM) (Jahangir et al., 2003b) although initial promising results were later withdrawn by the authors after discovering the HMM was effectively simulating a Gaussian Mixture Model (GMM) (Jahangir et al., 2003a). GMMs have been applied to the problem in their own right (Bilik et al., 2006) and shown capable of recognizing several classes through μ-DS analysis. This work also ignored the time variance of the signature and the durations of the input signals were much longer than the typical scanning radar dwell time. Through the application of image processing techniques it has been shown that the spectrogram of a μ-DS can be recognized using classifiers that do not ordinarily allow for time variance (Jiajin and Chao, 2005; Jiajin, 2005). However, the signatures used were simulated leading to spectrograms with substantially different natures to the real world examples presented in the other works cited here and as are presented in the following chapters of this document.

A broad appraisal of μ-DS based radar ATR research would be that it is an emerging field showing great potential. The requirements it places on the radar hardware are much lower than those associated with its main alternatives—HRRP, ISAR and SAR based ATR—since it only requires the radar to be coherent rather than have very high range resolution. These simple requirements suggests that it may be possible to implement μ-DS based radar ATR on older systems by updating the cheap digital signal processing rather than the expensive RF hardware.

One emerging area of radar ATR work that has yet to be investigated for μ-DS solutions is multiple

perspective classification. This concept has emerged from the domain of HRRP, ISAR and SAR ATR methods and uses multiple views of the target, separated in angle, to increase the amount of information passed to the classifier.

2.5.2.4 Multiperspective Radar ATR

Multiperspective radar ATR is a target recognition technique where multiple views of the target are combined to aid classifier performance. This increase in complexity of the classification approach is required when the output of the radar is highly dependent on the aspect angle of the target. In the case of HRRP of air targets it has been reported that, for an X-band radar, the aspect angle may only need change by a few tenths of 1° to cause a large change in the target's range profile (Liao et al., 2000). Techniques, such as averaging the HRRPs over a range of angles proposed by Liao, have been considered to reduce this sensitivity, but it has also been shown that simultaneously using multiple perspectives can improve the correct recognition rate.

The most common approach to simultaneously considering multiple perspectives is to employ a HMM. This method has been shown to improve the recognition rate by around 20% in applications such as recognizing air targets (Zhu et al., 2007) in ISAR images and distinguishing ground targets from trees in SAR images (Runkle et al., 2000). More modest results, showing gains of around 10%, have been reported for sonar based ATR when HMM and ANN are used together (Robinson et al., 2005). To date, the most comprehensive survey of multiperspective radar classification has been performed by Michele Vespe (Vespe et al., 2007; Vespe, 2007). This work contrasts the performance of three conventional classification systems—k-NN, naïve Bayesian and ANN—all of which have been modified to support the fusion of multiple perspectives. The performance gains vary between the different classifiers with the naïve Bayesian showing the largest increase in recognition rate of 27%. It is also shown that the more perspectives are included the greater the increase in recognition rate although most increase is achieved by moving from one to two perspectives. With such impressive performance gains possible it would seem pertinent to consider the possibility of applying multiperspective processing to μ-DS based radar ATR.

The final aspect of contemporary literature to be considered is that relating to the formal methodology of investigating radar ATR.

2.5.3 Limits To Current ATR Methodology

Regardless of the type of radar ATR being investigated there is a common limitation present in nearly all the open literature: there is no consistent method used for the investigations. This failure in consistency makes it difficult to contrast the results of different investigations particularly when the number and type of targets used differs between studies. With the exception of the recent conference paper (Smith et al., 2008a) based on Chapter 3 of this research and the work of Vespe (Vespe, 2007), the only guide to eval-

uating radar ATR performance is Marvin Cohen's tutorial notes from the 2006 IEEE Radar Conference (Cohen, 2006). In the notes Cohen introduces the idea that there are parameters other than the probability of correct classification, P_{cc}, that can be derived from the confusion matrix and provide improved evaluation of the classifier. Focus is given to how the reliability, RL, parameter may indicate poor classifier performance even in the presence of high P_{cc}. The notion that a practical classifier must have the option of declaring "unknown", to prevent forced declarations, is also considered. This leads on to a preliminary discussion of how Receiver Operating Characteristic (ROC) curves may be used to set the internal threshold required to prevent forced declarations. Unfortunately the nature of the document, tutorial notes, means the concepts are presented without justification, detailed analysis or references. The use of ROC curves to implement "unknown" declaration was extended into a powerful two curve approach in a conference paper concerned with multiperspective classification (Vespe et al., 2006a). This extension permitted further evaluation parameters, relating to the success of the classifier at correctly detecting "unknown" targets, to be defined but still does not result on a concrete methodology. For the most part researchers tend to rely on P_{cc} as the only evaluation metric of merit and the method in which a classifier is investigated varies greatly.

The lack of research into radar ATR evaluation was a key motivator for proposed evaluation methodology presented in Chapter 3 of this thesis.

2.6 Summary

This chapter introduced the fundamentals of radar that are of relevance to this research and précied the current scientific literature relating to μ-DS based radar ATR.

The fundamentals of radar were discussed using a block model approach. It was noted that the model used was more general than that used in many radar texts in order to facilitate the discussion of multistatic radar that will come in Chapter 8. The review included details of the ways radars calculate range, how they process Doppler information to detect motion and how the waveform may effect performance. Although brief, the review is sufficient enable the reader to appreciate the work within the thesis. Where necessary additional theory is presented *in situ*.

The literature survey outlined the current publications of interest to the presented work. Consideration was given to the current work on multistatic radars before turning to radar based ATR and classification. Initially, the publications relating to the classifiers used in the work were commented on. After this the various alternatives to μ-DS based radar ATR were summarized. The discussion of μ-DS radar ATR publications began by noting that JEM ATR is a precursor to μ-DS ATR. The various contemporary papers on the use of μ-DS in radar were then reported on. The consideration of radar ATR publications was concluded by looking at the use of multiple perspectives in classification. This is a technique that has shown great success in HRRP and ISAR based radar ATR methods and it was speculated that it will

be useful in μ-DS approaches also. The literature survey was closed with a discussion of the lack of formal methodology used in radar ATR. It was noted that this lack of methodology was a key motivator for research of Chapter 3.

Chapter 3

Evaluation Of Classification Systems

It was identified in section 2.5 that the study of radar Automatic Target Recognition (ATR) is not conducted in line with a formal methodology. This limitation creates difficulty in comparing the results of different studies, particularly when different datasets have been used. Here, a series of evaluation metrics are presented that provide a comprehensive analysis of the performance of an ATR system under test. Furthermore, a preliminary method by which the performance of a classifier may be predicted is proposed.

The chapter also provides a generic representation of a radar classification system, the "black box" classifier, against which the evaluation parameters are defined. As part of this generic representation features that all radar ATR solutions should exhibit are proposed.

Before embarking on an investigation of radar Automatic Target Recognition (ATR) it is necessary to consider how the various classification methods that might be used can be evaluated. Broadly speaking the problem of performing recognition can be broken into two stages: feature extraction and classification. There are several different approaches that may be taken for each and the evaluation process must be unaffected by the methods selected for each stage. To support this independence a "black box" classifier, that is a generalized model of the classification process, was developed. A series of evaluation metrics were then defined against the "black box" classifier and these are suitable for use assessing any practical implementation of a classifier. As part of this discussion so called "unknown" inputs, inputs that have no representation in the classifier's reference dataset, and variant inputs, that reflect variations of classes known to the classifier, will be discussed. The discussion of classification evaluation will close with consideration of how to predict performance.

The aim of performance prediction is to calculate expected values for the evaluation metrics. With such predictions made it becomes possible to consider the performance of a radar ATR solution against its theoretical best performance as well as other possible solutions. The subject was not considered in detail during the research and its importance only realized in the closing months of the project, as a result focus is given to the prediction of a single metric, the probability of correct classification.

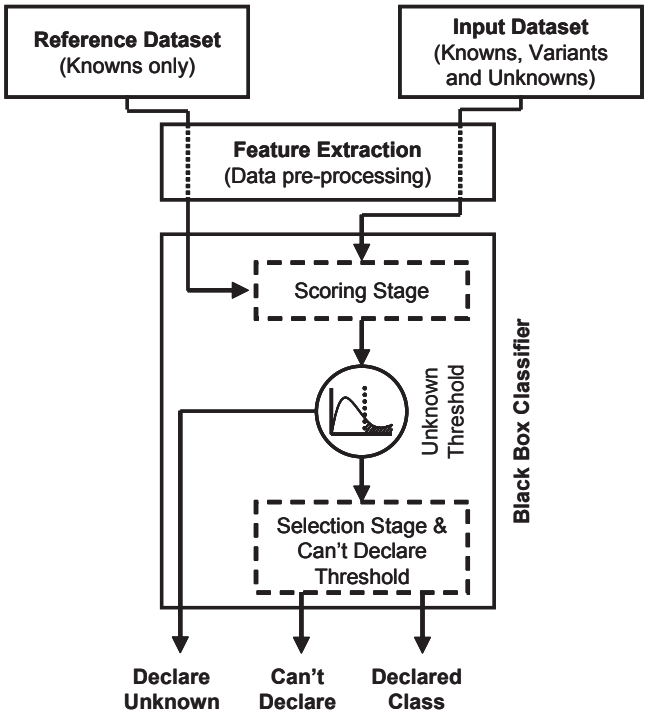


Figure 3.1: The "black box" classifier.

3.1 The Black Box Classifier

The "black box" classifier outlines the different parts of a generalized classification system, see Fig. 3.1. There are two sets of inputs: the reference dataset that contains pre-classified examples of input data and defines which classes are known to the classifier; and the input dataset that contains *all* the inputs the classifier may encounter during operation. The two datasets contain the raw radar output, the Doppler spectrum or High Range Resolution Profile (HRRP) for example, and require pre-processing before being passed to the classifier. The pre-processing forms the feature extraction stage in which details, such as the location of peaks, bandwidth of spectrum, etc. are extracted from the raw data into a feature vector for classification. Feature extraction is applied to both the reference and input datasets entries. A feature vector from the input dataset is passed into the "black box" classifier and into the scoring stage where it is assigned a series of scores indicating how similar it is to each class represented in the reference dataset. The scoring may take many forms (Duda et al., 2005): in parametric classifiers the reference dataset is used to estimate the parameters of class models prior to operation; similarly in learning algorithms the internal parameters of the classifier, such as the weights in an Artificial Neural Network (ANN), will have been found using the reference dataset in an offline training stage; while in a non parametric technique, such as k-Nearest Neighbour (k-NN), a direct comparison between the input and all entries of the reference dataset will be used for every classification performed. Regardless of the method used, a score for each of the known classes will be produced and these are passed onto the

"unknown" detection threshold.

It is not possible for the reference dataset to contain examples of every possible class. There will be omissions due to both lack of data and decisions not to include some classes because they are irrelevant to the situation in which the classifier will be deployed. However, failure to include a class in the reference dataset does not remove it from the input dataset and it may still be encountered during operation. When such a class is input to the classifier it is said to be of an "unknown" class and the classifier must be able to declare it as such rather than being forced to declare it as one of the known classes. The detection is made through application of a threshold to the scores emerging from the scoring stage. If the scores do not pass the threshold then the input is declared "unknown" and no further processing takes place. Clearly the level of the threshold is a key operational parameter of the classifier and setting it is an important part of the classifier development, see section 3.4.

Assuming that the input was not declared as "unknown" then a class declaration will be made based on the scores. During the selection stage, bottom of Fig. 3.1, the scores are assessed to see which class from the reference dataset is most likely to have resulted in the input feature vector. Typical approaches to making the selection include choosing the class which results in the largest/smallest score or performing a majority vote. Whichever approach is taken the objective is to select a single class and present it to the user. Unfortunately it is possible that the score of two or more classes are equal making the selection impossible. There are two possibilities for coping with such a situation: make an arbitrary decision on which class should be declared, based, for example, on an alpha-numeric sort of class name; or do not declare. The "can't declare" threshold is part of the selection stage and is intended to detect draws in the score for each class. Rather than fail to declare only in the event of a tie it is better to define a minimum separation between the winning score and the others. If the separation is insufficient then the classifier indicates that it can't declare. It is rare to find radar ATR systems that employ this final threshold: every system investigated in the literature review for this work did not include the "can't declare" threshold.

3.2 Inputs To The Classifier And "Unknown" Detection

The classes that are known to the classifier are defined by the content of the reference dataset. The set of known classes are given the label K and defined as:

$$K = \{k_1, k_2, k_3, \ldots k_{N_k}\} = \{k_n\} \quad \text{where} \quad \{n \in \mathbb{N} : n \leq N_k\} \tag{3.1}$$

where k_n is the class label for the n^{th} known class and N_k is the number of known classes. For each k_n there is a set of data series that the class causes the radar (or sensor in a broader case) to produce denoted

by F_{k_n} defined as:

$$F_{k_n} = \left\{ \phi_1^{(k_n)}, \phi_2^{(k_n)}, \phi_3^{(k_n)}, \ldots \right\} = \left\{ \phi_i^{(k_n)} \right\} \text{ where } \{i \in \mathbb{N}\} \tag{3.2}$$

where $\phi_i^{(k_n)}$ is the i^{th} data series of class k_n. Hypothetically i may range to infinity since a feature vector will depend on the conditions under which it is recorded and there are an unlimited number of conditions. Under practical constraints, however, i will be limited by the system used to gather the data series. For examples, the limits of range resolution may cause the radar to be insensitive to very small target movements while the number of bits available for storing the samples of the data series provides a finite, if very large, number of permutations. Despite this upper limit on i there will still be far too many possible series per class to include them all in the reference dataset. The content of the reference dataset, L, is then a subset of the super set formed by the union of all the F_{k_n} i.e.

$$L \subset \bigcup_{n=1}^{N_k} F_{k_n}. \tag{3.3}$$

There are then a limited number of feature vectors from each F_{k_n} included in L; if the number of examples for class k_n is denoted by $M_{k_n}^{(L)}$ then L may be more closely defined as:

$$L = \bigcup_{n=1}^{N_k} \bigcup_{j=1}^{M_{k_n}^{(L)}} \phi_j^{(k_n)}. \tag{3.4}$$

It needs to be noted that the $\phi_j^{(k_n)}$ do not represent contiguous entries from the set F_{k_n} and ideally should be spread through it.

There also exists a set of variant classes, V, that allow for slight variations to known classes, which should still be declared as one of the known classes. A variant may occur when a particular instance of a class has been modified in some way such as changing the weapon configuration on a fighter jet, adjusting the loadings on a tank or adding a backpack to a personnel target. Each variant of a class is described by a numerical index q such that $k_n^{(q)}$ represents the q^{th} variant of class k_n. The set V is then defined as:

$$V = \bigcup_{n=1}^{N_k} \bigcup_{q=1}^{Q_n} k_n^{(q)} \tag{3.5}$$

where Q_n is the number of variants of the n^{th} class. Just as each class k_n had a set of feature vectors $\{\phi_i^{(k_n)}\}$ so does each variant class. The sets of feature vectors are defined by:

$$F_{k_n^{(q)}} = \left\{ \phi_1^{(k_n^{(q)})}, \phi_2^{(k_n^{(q)})}, \phi_3^{(k_n^{(q)})}, \ldots \right\} = \{\phi_i^{(k_n^{(q)})}\} \text{ where } \{i \in \mathbb{N}\} \tag{3.6}$$

and where $\phi_i^{(k_n^{(q)})}$ is the i^{th} feature vector of variant class $k_n^{(q)}$.

The final set of classes to be considered is those that are unknown to the classifier. The unknown classes form the set U defined as:

$$\begin{aligned} U &= \{u_1, u_2, u_3, \ldots u_{N_u}\} = \{u_n\} \text{ where } \{n \in \mathbb{N}\} \\ U &\notin K \cup V \end{aligned} \tag{3.7}$$

where u_n is the class label for the n^{th} unknown class. As with the variant classes the size of U is potentially infinite: if the classes are unknown then it can not be known how many of them there are. Unknown radar targets may come from a wide range of real world sources. In a Synthetic Aperture Radar (SAR) image that is being analysed for military vehicles unknown targets may superficially be considered as types of tank or artillery that are not included in the set of known classes, K. Unfortunately there are many more possible candidates for unknowns than this since anything that can trigger a detection, such a tree, bolder, abandoned car or even a discarded refrigerator, will cause a valid data series to be passed to the feature extractor and on to the classifier. The data series of the unknown classes are defined in a similar manner to the known and variant classes:

$$F_{u_n} = \left\{\phi_1^{(u_n)}, \phi_2^{(u_n)}, \phi_3^{(u_n)}, \ldots\right\} = \{\phi_i^{(u_n)}\} \text{ where } \{i \in \mathbb{N}\} \tag{3.8}$$

and where F_{u_n} is the set of data series for the n^{th} unknown and $\phi_i^{(u_n)}$ is the i^{th} feature vector of unknown class u_n.

The input data set, I that contains all the possible inputs to the classifier, Fig. 3.1, is then the union of the known set, K, the variant set, V and the unknown set, U i.e.

$$I = K \cup V \cup U. \tag{3.9}$$

Although the classifier may encounter any input from the set I during operation it is not possible to obtain all the data series in I for use during testing since theoretically the set is of infinite size. Instead a representative subset of I must be collected to test the classifier. This dataset is the test dataset, T and is a subset of I:

$$T \subset I. \tag{3.10}$$

As for the reference dataset it is also possible to define the test dataset in terms of the data series it contains:

$$T = \bigcup_{N=1}^{N_k} \left\{ \left[\bigcup_{j=1}^{M_{k_n}^{(T)}} \phi_j^{(k_n)} \right] \cup \left[\bigcup_{q=1}^{Q_n} \bigcup_{j=1}^{M_{k_n^{(q)}}^{(T)}} \phi_j^{(k_n^{(q)})} \right] \right\} \cup \left\{ \bigcup_{n=1}^{N_u^{(T)}} \bigcup_{j=1}^{M_{u_n}^{(T)}} \phi_j^{(u_n)} \right\} \tag{3.11}$$

		Declared Class						
		k_1	k_2	k_3	k_4	...	k_{N_k}	Unknown
Input Class	k_1	d_{11}	d_{12}	d_{13}	d_{14}	...	d_{1N_k}	d_{1u}
	k_2	d_{21}	d_{22}	d_{23}	d_{24}	...	d_{2N_k}	d_{2u}
	k_3	d_{31}	d_{32}	d_{33}	d_{34}	...	d_{3N_k}	d_{3u}
	k_4	d_{41}	d_{42}	d_{43}	d_{44}	...	d_{4N_k}	d_{4u}
	$\vdots$					$\vdots$		
	k_{N_k}	d_{N_k1}	d_{N_k2}	d_{N_k3}	d_{N_k4}	...	$d_{N_kN_k}$	d_{N_ku}
	$k_1^{(1)}$	$d_{11}^{(1)}$	$d_{12}^{(1)}$	$d_{13}^{(1)}$	$d_{14}^{(1)}$	...	$d_{1N_k}^{(1)}$	$d_{1u}^{(1)}$
	$\vdots$					$\vdots$		
	$k_{N_k}^{(Q_n)}$	$d_{N_k1}^{(Q_n)}$	$d_{N_k2}^{(Q_n)}$	$d_{N_k3}^{(Q_n)}$	$d_{N_k4}^{(Q_n)}$	...	$d_{N_kN_k}^{(Q_n)}$	$d_{N_ku}^{(Q_n)}$
	u_1	$d_{11}^{(U)}$	$d_{12}^{(U)}$	$d_{13}^{(U)}$	$d_{14}^{(U)}$	...	$d_{1N_k}^{(U)}$	$d_{1u}^{(U)}$
	$\vdots$					$\vdots$		
	u_{N_u}	$d_{N_u1}^{(U)}$	$d_{N_u2}^{(U)}$	$d_{N_u3}^{(U)}$	$d_{N_u4}^{(U)}$	...	$d_{N_uN_k}^{(U)}$	$d_{N_uu}^{(U)}$

Table 3.1: A prototype declaration matrix.

where $M_{k_n}^{(T)}$ is the number of feature vectors of the class k_n in the test dataset; $M_{k_n^{(q)}}^{(T)}$ the number of feature vectors of the variant class $k_n^{(q)}$ in the test dataset; and $M_{u_n}^{(T)}$ the number of feature vectors of the unknown class u_n in the test dataset. Again, as in the definition of the reference dataset the feature vectors enumerated by the index j are not intended to be contiguous entries from the sets F_x where $x = \{k_n, k_n^{(q)}, u_n\}$, but rather a representative spread through each set.

3.3 Evaluation Metrics

It may seem that when evaluating a classifier the only question of interest is, "How often does the classifier declare the correct class?" The situation is more involved than this, however, and a number of metrics are required to assess performance. Typically the metrics of interest take the form of probabilities, or likelihoods, and help answer questions such as, "What is the likelihood that declared class is the same as the input class?" The values of these probabilities are found during a testing stage when the test dataset, T, that was described in section 3.2 is applied to the classifier. It is assumed that the tester knows the correct class of each input and is therefore able to keep a detailed log of the classifier output in a declaration matrix, Table 3.1. The declaration matrix consists of three types of element. The first, d_{ij}, counts the number of outputs for the class k_j when data from the class k_i is presented to the classifier; the second $d_{ij}^{(q)}$ counts the outputs for class k_j when data from the variant class $k_i^{(q)}$ is presented, as before q is used to index the variants for a particular known class; and lastly $d_{ij}^{(U)}$ counts declarations for k_j when unknown class u_i is presented. In each type of elements when $j = u$ the count is for declarations of "unknown". Through analysis of the declaration matrix it is possible to calculate the declaration

likelihoods of particular classes.

The probability that the classifier declares a particular class depends on the input class and is given the symbol P_{ij}. It is defined as:

$$P_{ij} = \mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_i) \tag{3.12}$$

that is: the probability of making a declaration for class k_j when class k_i is presented. It may be calculated from the declaration matrix according to:

$$P_{ij} = \frac{d_{ij}}{M_{k_i}^{(T)}}. \tag{3.13}$$

where d_{ij} and $M_{k_i}^{(T)}$ are the number of declarations made for class k_j when class k_i is presented and the number of entries for class k_i in the test dataset respectively (see Table 3.1 in section 3.2).

The case when $i = j$ in (3.13) leads to the first evaluation metric: the *probability of correct classification*. The probability of correct classification is denoted by $P_{\text{cc}}(k_n)$, and it is the probability of correctly identifying the class k_n. Typically $P_{\text{cc}}(k_n)$ is only considered for the known classes although it is possible to calculate it for both variants and unknowns by considering, respectively, $k_n^{(q)} = k_n$ and a declaration of "unknown" for any class u_n as being correct. There is a unique value of $P_{\text{cc}}(k_n)$ for each class; the classifier's overall probability of correct classification is the average:

$$P_{\text{cc}} = \langle P_{\text{cc}}(x) \rangle_{x \in K}\,. \tag{3.14}$$

Many works on radar ATR only consider probability of correct classification in the evaluation of the classifier, however it is not able to provide a full description of the classifier's performance. For instance, it does not answer the question, "What is the probability that the input feature vector came from class k_i if the classifier declared class k_j?" For the case where $i \neq j$ this question may seem purely academic in nature, but when $i = j$ it is of utmost importance since it provides an indication of whether the classifier output may be relied upon. Formally the answer to the question is the posterior probability $\mathrm{P}(\text{I/P} = k_i \,|\, \text{Dec} = k_j)$ that can be found from (3.12) through the application of Bayes's Theorem (Duda et al., 2005):

$$\mathrm{P}(\text{I/P} = k_i \,|\, \text{Dec} = k_j) = \frac{\mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_i)\mathrm{P}(k_i)}{\sum_{x \in T} \mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_x)\mathrm{P}(k_x)}. \tag{3.15}$$

In (3.15) $\mathrm{P}(k_i)$ is the prior probability of encountering an input from class k_i; $\mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_i)$ the likelihood of declaring class k_j, as defined above, and $\sum_{x \in T} \mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_x)\mathrm{P}(k_x)$ is a normalizing constant that reduces to the probability of declaring class k_j, $\mathrm{P}(\text{Dec} = k_j)$, independent of

the input class. The prior probability can be found according to:

$$\mathrm{P}(k_i) \quad = \quad \frac{M_{k_i}^{(T)}}{\sum_{j=1}^{N_k} M_{k_j}^{(T)}} \tag{3.16}$$

where the $M_{k_j}^{(T)}$ are the number of data series included in the test dataset for each class from (3.11). The likelihood expression is the same as P_{ij} defined in (3.12) and can be calculated from (3.13). The normalizing constant is then found by summing the various products of likelihood and prior probability as described in (3.15). Unfortunately, when calculated in this manner the posterior probability is dependent on the test dataset used and does not give an objective insight into the classifier's performance.

The posterior probability depends on the prior probability of encountering a particular class that was calculated above from the number of entries in the test dataset; it is not valid, however, to assume that the prior probability for the test dataset is the same as the true prior probability. For radar ATR the true prior probability for a target class will depend on the number of them in the region in which the radar is being deployed relative to the number of all other targets. There is no reason to assume that the distribution of targets in the test dataset will match the distribution in a real environment. Rather than attempting to make estimates of the prior probability it has been suggested (Cohen, 2006) that it be assumed equal for all classes; this assumption allows for the definition of the second evaluation parameter, *reliability*, that is given the symbol $RL(k_n)$. If the now equal prior probabilities $\mathrm{P}(k_x)$ are given the constant symbol P_{k_n} then from (3.15):

$$\begin{aligned}
\mathrm{P}(\text{I/P} = k_i \,|\, \text{Dec} = k_j) \quad &= \quad \frac{\mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_i) P_{k_n}}{\sum_{x \in K} \mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_x) P_{k_n}} \\
&= \quad \frac{P_{k_n}}{P_{k_n}} . \frac{\mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_i)}{\sum_{x \in K} \mathrm{P}(\text{Dec} = k_j \,|\, \text{I/P} = k_x)} \\
&= \quad \frac{P_{ij}}{\sum_{j=1}^{N_k} P_{ij}} \\
&= \quad RL(k_i) \;\text{ if }\; i = j.
\end{aligned} \tag{3.17}$$

As with the probability of correct classification, reliability is defined for each of the known classes. To find a general classifier level parameter an average can be taken:

$$RL = \langle RL(x) \rangle_{x \in K} . \tag{3.18}$$

The two evaluation metrics defined so far have been limited to providing information about the known classes only. Although it is possible to extend them to include the variant classes these are given their

own evaluation metric.

The *probability of generalization*, or *generalization* for short, is the probability that when a variant class, $k_n^{(q)}$, is presented it is declared as the correct known class, k_n. It may be calculated as:

$$P_{\text{Gen}}(k_i) = \left\langle \text{P}(\text{Dec} = k_i \,|\, \text{I/P} = k_i^{(q)}) \right\rangle_q \tag{3.19}$$

$$\approx \left\langle \frac{d_{ij}^{(q)}}{M_{k_n^{(q)}}^{(T)}} \right\rangle_q \quad \text{where} \quad \{q \in \mathbb{N} \,:\, q < Q_i\} \tag{3.20}$$

which is the average generalization across all of the variants of a particular known class. Equation (3.19) provides the true $P_{\text{Gen}}(k_i)$ and (3.20) the estimate based on T. As for the other metrics it too may be averaged to give a classifier level indication of the variance:

$$P_{\text{Gen}} = \langle P_{\text{Gen}}(x) \rangle_{x \in K} \,. \tag{3.21}$$

The final two metrics relate to the "unknown" threshold. The first describes the rate with which known targets correctly pass the threshold and the second, the rate with which unknown targets incorrectly pass it. Assessment of these metrics is essential for deciding on the level for the "unknown" threshold. The rate at which known classes pass the threshold is the *probability of declaration*. The metric is defined as:

$$P_{\text{Dec}}(k_i) = \sum_{j=1}^{N_k} P_{ij} = 1 - P_{iu} \tag{3.22}$$

where P_{iu} an extension to (3.12) to indicate the probability of declaring unknown when class k_i is presented to the classifier. The overall probability of declaration is the average over all the known classes:

$$P_{\text{Dec}} = \langle P_{\text{Dec}}(x) \rangle_{x \in K} \,. \tag{3.23}$$

The final metric is the *probability of false alarm*, the opposite of the probability of declaration. When an unknown input is incorrectly declared as one of the knowns, then a false alarm is said to have occurred. For any of the known classes the probability of false alarm $P_{\text{FA}}(k_n)$ is the average of the probability of declaring k_n when an unknown class is presented:

$$P_{\text{FA}}(k_j) = \langle \text{P}(\text{Dec} = k_j \,|\, \text{I/P} = u_i) \rangle_{u_i \in U} \tag{3.24}$$

$$\approx \left\langle \frac{d_{ij}^{(U)}}{M_{u_i}^{(T)}} \right\rangle \quad \text{where} \quad \left\{ i \in \mathbb{N} \,:\, i \leq N_u^{(T)} \right\}. \tag{3.25}$$

As for $P_{\text{Gen}}(k_i)$, (3.24) is the true P_{FA} and (3.25) an estimate made from T. The probability of false

alarm can also be averaged over the unknown classes to provide a single metrics for the classifier:

$$P_{\mathrm{FA}} = \langle P_{\mathrm{FA}}(x) \rangle_{x \in U} \,. \tag{3.26}$$

Although it does not show the value of all of the metrics the confusion matrix is one of the most common methods for reporting the performance of a classifier. The basic confusion matrix (Duda et al., 2005) shows the likelihood of declaration, P_{ij}, for the known classes but it can be extended to include the reliability and provide enough information to calculate P_{Gen}, P_{Dec} and P_{FA} (Vespe, 2007; Smith et al., 2008a). A prototype confusion matrix is shown in Table 3.2 its elements relate the structure of the matrix back to the formulae for the evaluation metrics and the definition of the various datasets. Each row of the confusion matrix corresponds to a class from the test dataset, T. The rows are titled with the input class label and the number of data series entries that were available for that class. The columns of the matrix correspond to the possible declarations that the classifier may make. The first N_k columns are for the known classes and the final column for the "unknown" declaration. Each entry of the matrix is a declaration likelihood, P_{ij}, as defined in (3.12) and (3.13). (The concept of declaration likelihood has been extended to cover the variant and unknown classes with the same labelling scheme as the declaration matrix, Table 3.1.) Each row of the confusion matrix represents the distribution of declarations for a particular input class and as such should sum to 1.0. To ease more detailed explanation the confusion matrix has been broken down into three sections indicated to the right of the table:

- The *Test Target Matrix* details the results for the known target classes. The leading diagonal of this sub-matrix shows the probability of correct classification for the class, $P_{\mathrm{cc}}(k_n)$ that is equal to P_{ij} when $i = j$. Also shown as a bracketed superscript in the cells of the leading diagonal are the class reliabilities, $RL(k_n)$. Although it is possible to calculate reliabilities for the other elements of the test target matrix the values provide no significant insight to classifier performance so are omitted for clarity.
- The *Variant Target Matrix* details the performance of the variant classes. Each element is an extension to (3.12) and (3.13) that extends the calculation of P_{ij} to allow for variants:

$$P_{ij}^{(q)} = \frac{d_{ij}^{(q)}}{M_{k_i^{(q)}}^{(T)}}. \tag{3.27}$$

  Unlike the test target matrix it is not possible to read the key statistic, probability of generalization, from the table since it requires an average to be taken of the Q_n variants of the known class k_n.
- The *Unknown Target Matrix* details the declaration probabilities for unknown classes. The average of each column, bar the "unknown" declaration column, provides the probability of false

Input Class	Declared Class						
	k_1	k_2	k_3	k_4	$\cdots$	k_{N_k}	Unknown
Test Target Matrix							
$k_1\ (M_{k_1}^{(T)})$	$P_{\text{cc}}\left(k_1\right)^{RL(k_1)}$	P_{12}	P_{13}	P_{14}	$\cdots$	P_{1N_k}	P_{1u}
$k_2\ (M_{k_2}^{(T)})$	P_{21}	$P_{\text{cc}}\left(k_2\right)^{RL(k_2)}$	P_{23}	P_{24}	$\cdots$	P_{2N_k}	P_{2u}
$k_3\ (M_{k_3}^{(T)})$	P_{31}	P_{32}	$P_{\text{cc}}\left(k_3\right)^{RL(k_3)}$	P_{34}	$\cdots$	P_{3N_k}	P_{3u}
$k_4\ (M_{k_4}^{(T)})$	P_{41}	P_{42}	P_{43}	$P_{\text{cc}}\left(k_4\right)^{RL(k_4)}$	$\cdots$	P_{4N_k}	P_{4u}
$\cdots$					$\cdots$		
$k_{N_k}\ (M_{k_5}^{(T)})$	P_{N_k1}	P_{N_k2}	P_{N_k3}	P_{N_k4}	$\cdots$	$P_{\text{cc}}\left(k_{N_k}\right)^{RL(k_{N_k})}$	P_{N_ku}
Variant Target Matrix							
$k_1^{(1)}\ (M_{k_1^{(1)}}^{(T)})$	$P_{11}^{(1)}$	$P_{12}^{(1)}$	$P_{13}^{(1)}$	$P_{14}^{(1)}$	$\cdots$	$P_{1N_k}^{(1)}$	$P_{1u}^{(1)}$
$k_1^{(2)}\ (M_{k_1^{(2)}}^{(T)})$	$P_{11}^{(2)}$	$P_{12}^{(2)}$	$P_{13}^{(2)}$	$P_{14}^{(2)}$	$\cdots$	$P_{1N_k}^{(2)}$	$P_{1u}^{(2)}$
$\cdots$					$\cdots$		
$k_1^{(Q_1)}\ (M_{k_1^{(Q_1)}}^{(T)})$	$P_{11}^{(Q_1)}$	$P_{12}^{(Q_1)}$	$P_{13}^{(Q_1)}$	$P_{14}^{(Q_1)}$	$\cdots$	$P_{1N_k}^{(Q_1)}$	$P_{1u}^{(Q_1)}$
$\cdots$					$\cdots$		
$k_{N_k}^{(Q_n)}\ (M_{k_{N_k}^{(Q_n)}}^{(T)})$	$P_{N_k1}^{(Q_n)}$	$P_{N_k2}^{(Q_n)}$	$P_{N_k3}^{(Q_n)}$	$P_{N_k4}^{(Q_n)}$	$\cdots$	$P_{N_kN_k}^{(Q_n)}$	$P_{N_ku}^{(Q_n)}$
Unknown Target Matrix							
$u_1\ (M_{u_1}^{(T)})$	$P_{11}^{(U)}$	$P_{12}^{(U)}$	$P_{13}^{(U)}$	$P_{14}^{(U)}$	$\cdots$	$P_{1N_k}^{(U)}$	$P_{1u}^{(U)}$
$u_2\ (M_{u_2}^{(T)})$	$P_{21}^{(U)}$	$P_{22}^{(U)}$	$P_{23}^{(U)}$	$P_{34}^{(U)}$	$\cdots$	$P_{2N_k}^{(U)}$	$P_{2u}^{(U)}$
$u_3\ (M_{u_3}^{(T)})$	$P_{31}^{(U)}$	$P_{32}^{(U)}$	$P_{33}^{(U)}$	$P_{34}^{(U)}$	$\cdots$	$P_{3N_k}^{(U)}$	$P_{3u}^{(U)}$
$u_4\ (M_{u_4}^{(T)})$	$P_{41}^{(U)}$	$P_{42}^{(U)}$	$P_{43}^{(U)}$	$P_{44}^{(U)}$	$\cdots$	$P_{4N_k}^{(U)}$	$P_{4u}^{(U)}$
$\cdots$					$\cdots$		
$u_{N_u}\ (M_{N_u}^{(T)})$	$P_{N_u1}^{(U)}$	$P_{N_u2}^{(U)}$	$P_{N_u3}^{(U)}$	$P_{N_u4}^{(U)}$	$\cdots$	$P_{N_uN_k}^{(U)}$	$P_{N_uu}^{(U)}$

Table 3.2: A prototype confusion matrix.

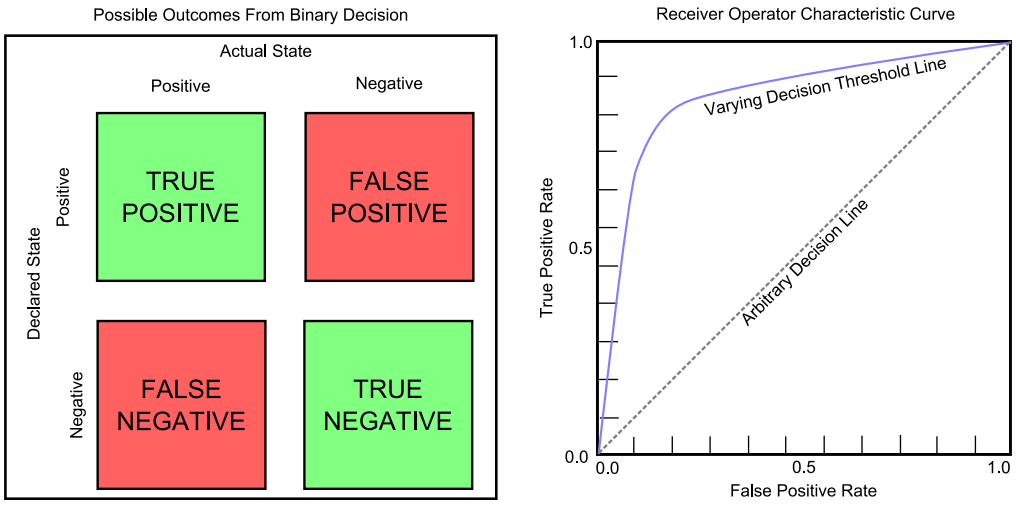


Figure 3.2: The output of a binary classifier (left) and how varying the internal threshold affects the true-positive and false-positive rates (right).

alarm for the relevant declaration class. Under ideal circumstances all entries of the unknown target matrix should be 0.0, bar those of the "unknown" declaration column that should be 1.0.

3.4 ROC Curves

The setting of the unknown class detection threshold greatly effects classifier performance: if it is too severe, known classes will inaccurately be declared as "unknown" causing the probability of declaration, P_{Dec}, and probability of correct classification, P_{cc}, to fall; but if it is too lax then many "unknowns" will be declared as known classes and the probability of false alarm, P_{FA}, will rise. Balancing of the threshold can be achieved through the use of Receiver Operating Characteristic (ROC) curves.

The ROC curve is a graphical device to aid in understanding of binary decisions and relates the rate of true positives to rate of false positives (Fawcett, 2003). It was originally developed to assist in radar signal detection during the Second World War (Egan, 1975) but is now used in a variety of applications ranging from psychology to artificial intelligence. In a binary decision a declaration is made for either positive or negative (in radar positive indicates a signal as present) leading to four possible results, left side Fig. 3.2. If a positive declaration is correctly made the result is a true-positive while if the positive declaration is incorrect the result is a false-positive. There are similar results for negative declarations. Typically the decision for a positive output is the result of an input score or level exceeding a threshold. In radar signal processing it is the amplitude of the signal exceeding a threshold that causes a positive declaration. But there is always a risk that the noise signal, section 2.2, will result in a positive declaration being made even when no signal is present (a false-positive result).

By varying the level of the signal detection threshold the probability of true-positives and false-positives are changed. The ROC curve plots the true-positive rate as a function of the false-positive

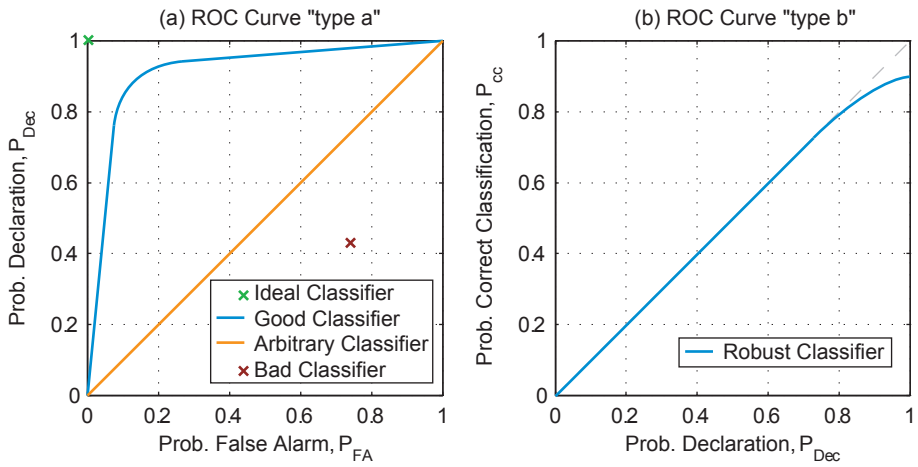


Figure 3.3: Example ROC curves.

rate, right side Fig. 3.2. Varying the threshold selects different points in the ROC space and permits the plotting of the blue line. When the threshold is lax every signal is detected since its amplitude always exceeds the threshold but the noise signal too always exceeds the threshold. This defines the top right limit of the ROC space when both rates are 1.0. At the other extreme the threshold is overly strict so neither noise nor signal can cause a positive output and both rates are equal to 0.0 defining the bottom left corner of the space. The threshold's optimum value causes the point on the top left corner of the blue curve: here the best balance between true- and false-positive rates is obtained. The ROC curve can be used for any binary decision, however, and not just radar signal processing.

Typically the decision making algorithm or classifier would assign a score to a particular input, see section 3.1, and then a threshold would be used to decide if the score was high enough to warrant a positive declaration being made. If the selection method is arbitrary then there is a 50–50 chance of the output being a true- or false-positive. Such a classifier leads to the arbitrary decision line, sometimes call the chance line, that runs through the forward diagonal of the ROC space (Egan, 1975), the dashed line on right side of Fig. 3.2. Although the classifier evaluation theory that has developed in this chapter has not been binary, the detection threshold for the unknown classes is a binary decision and ROC curves can be of great value is setting the threshold.

For the binary decision made by the "unknown" threshold of the "black box" classifier a declaration for a known class is considered a positive; therefore the probability of declaration is the true-positive rate and the probability of false alarm the false-positive rate. A plot of P_{Dec} as a function of P_{FA} is then a ROC curve, as introduced above, for the "unknown" detection process, left side Fig. 3.3. This curve is known as the "type a" curve to distinguish it from a second, "type b", curve that is required to assess the impact of the threshold on the probability of correct classification. The stricter the threshold the more

chance there is of known classes being declared unknown; this lowers the value of P_{cc} for the classifier. While it is possible to add a third axis to the "type a" curve to show the variation in P_{cc} it is clearer to plot it separately. The graph of P_{cc} as a function of P_{Dec} provides the second ROC curve, right side Fig. 3.3. Analysis of these two curves permits the level of the "unknown" threshold to be set in a quantifiable way.

An ideal radar ATR system would have $P_{FA} = 0$ and $P_{Dec} = 1$ and so occupy a point to the top-left of the "type a" ROC curve, the green cross in Fig. 3.3. A practical implementation, however, will not be ideal and varying the level of the "unknown" threshold will trace out a ROC curve that will vary between the point (0, 0) and (1, 1). For the system to be performing better than random selection of unknowns the curve must be above the line produced by an arbitrary classifier, the blue and orange lines Fig. 3.3. A poorly performing classifier would have a position below the line, the red cross in Fig. 3.3, indicating that it has more false alarms than known declarations. Once the threshold has been selected from the "type a" curve its probability of correct classification can be found from the "type b" curve. The ideal position on the "type b" curve is the top-right where $P_{Dec} = 1$ and $P_{cc} = 1$ but as before a practical implementation will be worse than this. The maximum value that P_{cc} can obtain is that of the P_{Dec} since a $P_{Dec} < 1$ means some known are being incorrectly declared as unknown. The "type b" ROC curve of a robust classifier therefore follows the leading diagonal of the space for the overly strict threshold and drops away from the diagonal when the threshold becomes lax and it is solely the classifier's class separation that is driving P_{cc}, blue line right side Fig. 3.3

3.5 The "Can't Declare" Threshold

The last element of the "black box" classifier to consider is the "can't declare" threshold that prevents the classifier making arbitrary decisions when there is no clear separation between potential declaration classes. Confusion between classes is not an indication that the classifier is not working properly. It is entirely possible that a feature vector could have resulted from any one of the known classes meaning it is not possible to make a definite declaration. Under such a situation the scores output by the classifier would be equal or very close. The "can't declare" threshold compares the leading scores to see if there is enough separation to make a definite declaration. If there is not, the classifier may either not declare; declare with an indication that the scores were close and that some heuristic has been used to decide which declaration to make; or declare all the possible known classes. Which ever option is taken the "can't declare" threshold will result in a reduction in probability of correct classification and probability of declaration (as they were defined in section 3.3).

By providing the system with a mechanism to not declare the values of P_{Dec} and P_{cc} will fall. The reduction in declaration performance, however, should result in an improvement in reliability. The reliability of a given known class, $RL(k_n)$ is reduced when other classes k_m, $m \neq n$, are mistakenly

declared as it since this causes an increase in the denominator term, $\sum_{j=1}^{N_k} P_{ij}$, from (3.17). Since the "can't declare" threshold prevents the chance of incorrect declaration in these situations it increases the value of RL at the expense of P_{Dec} and P_{cc}. Despite its potential improvements to the radar ATR system performance it was decided not to investigate the "can't declare" threshold as part of this work.

The use of the additional evaluation metrics and the "unknown" threshold described above require more test data to be available than if only P_{cc} in a forced declaration environment is used to evaluate a classifier. Including the "can't declare" threshold would require even more test data since it will reduce the number of test inputs that result in a declaration. It was decided that there was insufficient data available for this to be acceptable. Research into the effects of the "can't declare" threshold was therefore left for a future investigation.

3.6 Performance Prediction

The objective of performance prediction is to assess how well a classifier should perform given the data it is to classify. A full performance prediction would include making an estimate for all of the evaluation metrics that were defined in section 3.3 above. Unfortunately the significance of performance prediction was only realized during later stages of the research and there was insufficient time to develop methods to estimate all the metrics. Instead focus was give to the probability of correct classification, P_{cc}, and performance is estimated based on this metric only. Although it would have been preferable to estimate all of the evaluation metrics, the prediction of P_{cc} still represents a substantial advance in radar ATR methodology. Performance prediction is not generally addressed by the radar ATR community and was not discussed at any of the major radar conferences for 2006 and 2007[1] despite several session dedicated to ATR. Before moving to introduce the prediction method used in parts of this work, the prediction concept will be described in more detail.

If a binary classifier is considered that correctly recognizes the class of the input data 80% of the time, $P_{\text{cc}} = 0.8$, it is difficult to understand if the classifier is any good or not. The feature vectors that are fed into the classifier can be regarded as representing points in the feature space. The inputs from the two classes form clouds of points in the space and regions of the space that contain a particular class may be defined. If there is no overlap between the regions for each class then the data are completely separable and it should be possible to correctly classify inputs 100% of the time, $P_{\text{cc}}^{\max} = 1.0$. It is assumed that this level of data separation is achievable for the case considered so far. In a separate study a different binary classifier recognizes classes from a different dataset just 9% of the time, $P_{\text{cc}} = 0.09$. Is this classifier worse than the first? Regarding just the headline figure for P_{cc} may suggest that it is. However, if analysis of the regions in the second feature space show the two classes to overlap considerably and it

[1] EuRAD 2006, Manchester; CIE International Conference On Radar 2006, Shanghai; IET Conference On Radar Systems 2007, Edinburgh

is determined that it is only possible to correctly identify the classes 10% of the time,$P_{\text{cc}}^{\max} = 0.1$ then the second classifier may be regarded as having the best performance since 9% is a higher percentage of 10% than 80% is of 100%. Performance prediction is an attempt to estimate the theoretical maximum of an evaluation metric to see how well the classifier is expected to perform. The estimate is made using the data that has passed through the feature extraction stage of the "black-box" classifier, Fig. 3.1, since the data before this stage is not in the feature space the classifier operates on.

Although the intention was to estimate the metric for probability of correct classification the approach taken investigated the probability of error, $\mathrm{P}(\epsilon)$, that is equal to 1-P_{cc}. The approach was selected as there are known simplifications for calculating $\mathrm{P}(\epsilon)$ (Kailath, 1967; Duda et al., 2005) and its direct relationship to P_{cc} makes it just as valuable. When classifying between multiple classes $\mathrm{P}(\epsilon)$ can be defined in terms of P_{cc}:

$$\begin{aligned} \mathrm{P}(\epsilon) &= 1 - P_{\text{cc}} \\ &= 1 - \sum_{i=1}^{N_c} \mathrm{P}(c_i) \int_{\chi_i} p(\mathbf{x}|c_i) \mathrm{d}\mathbf{x} \end{aligned} \tag{3.28}$$

where N_c is the number of classes, χ_i is the region of feature space defined by the classifier's decision boundaries as containing samples of class c_i, $\mathbf{x}$ is a feature vector, and $\mathrm{P}(c_i)$ is the prior probability of observing class c_i. If there are only two classes then $\mathrm{P}(\epsilon)$ may be expressed directly in terms of the probability of error for each class:

$$\mathrm{P}(\epsilon) = \mathrm{P}(c_1) \int_{\chi_2} p(\mathbf{x}|c_1) \mathrm{d}\mathbf{x} + \mathrm{P}(c_2) \int_{\chi_1} p(\mathbf{x}|c_2) \mathrm{d}\mathbf{x} \tag{3.29}$$

i.e. the sum of the probabilities of obtaining a feature vector in the "wrong" region of feature space for each class. If the Bayes decision rule is used, the classifier's decision boundaries result in (3.28) and (3.29) gives the optimal value for $\mathrm{P}(\epsilon)$ (Duda et al., 2005). Unfortunately, however, since χ is a multidimensional space, performing the necessary integrals is very difficult even using numerical techniques.

Although the integrals are intractable (Duda et al., 2005), it is possible to determine the upper bound of $\mathrm{P}(\epsilon)$ for the two class case when the feature vectors for each class have a normal distribution as (Duda et al., 2005):

$$\mathrm{P}(\epsilon) \leq \mathrm{P}(c_1)^{\beta} \, \mathrm{P}(c_2)^{1-\beta} \int_{\chi} p^{\beta}(\mathbf{x}|c_1) p^{1-\beta}(\mathbf{x}|c_2) \mathrm{d}\mathbf{x} \quad \{\beta \in \mathbb{R} \,:\, 0.0 \leq \beta \leq 1.0\}. \tag{3.30}$$

If the probability distributions are normal, then the integral may be analytically evaluated as:

$$\int_{\chi} p^{\beta}(\mathbf{x}|c_1) p^{1-\beta}(\mathbf{x}|c_2) \mathrm{d}\mathbf{x} = e^{-k(\beta)} \tag{3.31}$$

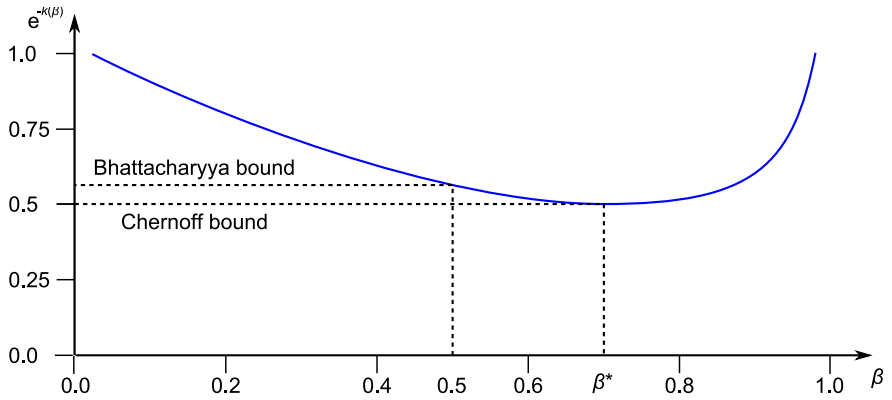


Figure 3.4: The Chernoff and Bhattacharyya bounds that lead to the lower limit of $\mathrm{P}(\epsilon)$.

where

$$\begin{aligned} k(\beta) &= \tfrac{\beta(1-\beta)}{2}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^{\mathrm{T}}\left[\beta\boldsymbol{\Sigma}_1+(1-\beta)\boldsymbol{\Sigma}_2\right]^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1) \\ &\quad +\tfrac{1}{2}\ln\frac{|\boldsymbol{\beta\Sigma}_1+(1-\beta)\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|^{\beta}|\boldsymbol{\Sigma}_2|^{1-\beta}} \end{aligned} \tag{3.32}$$

A graph showing a representative plot of (3.32) is shown in Fig. 3.4. By finding the β^* that minimizes the curve the value of the integral in (3.31) can be calculated leading to a value for the upper bound on $\mathrm{P}(\epsilon)$ from (3.30). Using β^* leads to finding the Chernoff bound on the error which is much simpler than finding the true error since minimizing $e^{-k(\beta)}$ only requires optimization in a one dimensional space.

Simpler still is to use the Bhattacharyya bound on the error in which β is assumed to be $\frac{1}{2}$ leading to:

$$\begin{aligned} \mathrm{P}(\epsilon) &\leq \sqrt{\mathrm{P}(c_1)\,\mathrm{P}(c_2)}\int_{\chi}\sqrt{p(\mathbf{x}|c_1)p(\mathbf{x}|c_2)}\mathrm{d}\mathbf{x} \\ &\leq \sqrt{\mathrm{P}(c_1)\,\mathrm{P}(c_2)}e^{-k(1/2)} \end{aligned} \tag{3.33}$$

where, for the normal distribution,

$$\begin{aligned} k(1/2) &= \tfrac{1}{8}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^{\mathrm{T}}\left[\frac{\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2}{2}\right]^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1) \\ &\quad +\tfrac{1}{2}\ln\frac{\left|\frac{\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2}{2}\right|}{\sqrt{|\boldsymbol{\Sigma}_1||\boldsymbol{\Sigma}_2|}} \end{aligned} \tag{3.34}$$

and $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the means and covariance matrices for the two classes. The Bhattacharyya bound is also shown in Fig. 3.4 where it is seen that it provides a slightly higher value for $e^{-k(\beta)}$ that results in a larger valued estimate of $\mathrm{P}(\epsilon)$.

The derivation of the Bhattacharyya bound on the error assumed that there are two classes to be discriminated between, but in practice there could be many more. A solution to this is to consider the error between pairs of targets (Smith et al., 2008c). While a multi-class classifier will not provide output of the probability of discriminating one class from another, it is a trivial exercise to calculate this information from the declaration and confusion matrices. There are, however, further limits on use of the Bhattacharyya bound as a performance predictor.

Firstly, in the calculation of the error bound the prior probabilities of the classes, $\mathrm{P}(c_n)$, are required. It is unlikely that these will be known in a real situation and it is not practical to estimate them from the reference dataset (Smith et al., 2008a). To enable the estimator to be developed for this work the probabilities were assumed equal and set to 0.5 in all calculations of the errors between pairs. Secondly, the form of the probability density functions, $p(\mathbf{x}|c)$, will not be known for real radar data and it is highly unlikely that it would have a normal distribution. When the error bound was used in this research this discrepancy was ignored and the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ were calculated as if the data was normally distributed. And thirdly, the bound on $\mathrm{P}(\epsilon)$ does not allow for the action of the "unknown" threshold. However, since the intention of the threshold is to discard targets that are not in the reference dataset it was felt it would have limited impact on the estimate of $\mathrm{P}(\epsilon)$ since these targets are not used in calculating the bound. The most likely effect of the threshold is to increase $\mathrm{P}(\epsilon)$ as known targets mistakenly declared as unknown reduce the value of P_{cc} and $\mathrm{P}(\epsilon) = 1 - P_{\mathrm{cc}}$.

3.7 Summary

The basis of a methodology for evaluating radar ATR solutions, and a candidate method for predicting their performance prior to implementation, has been outlined. To support these concepts a "black box" model of a radar ATR system was presented that contained the key stages required to perform recognition but did not specify their implementations. From this model a set of five evaluation metrics were defined.

The probability of correct classification, P_{cc}, was defined. It describes the probability that a classifier correctly identifies targets represented in its reference dataset and is the most common of all the evaluation parameters. Associated to with P_{cc} was the reliability, RL, that is the probability that the class input to the system was the same as the class it declared. The next two parameters relate to the system's ability to detect inputs that come from so called unknown classes that are not represented in the reference dataset. The probability of false alarm, P_{FA}, is the probability that an "unknown" input is declared as one of the know classes while the probability of declaration, P_{Dec}, is probability that the known classes are declared at all. Between them these parameters form a true-positive/false-positive pairing, such as is used in radar detection theory, allowing ROC analysis to be used to evaluate the level of the "unknown" threshold. The last parameter to be proposed was the probability of generalization, P_{Gen}, that describes the ability of the classifier to generalize or extrapolate from its reference data. Generalization is the ability to correctly identify variants to the known classes. With these five parameters it was considered that a comprehensive evaluation of the classifier's performance could be made. The close of the chapter then presented a method by which the classifier's performance could be predicted prior to implementation.

Predicting the performance of a radar ATR system ahead of implementation allows better assessment of its behaviour. If the nature of the data to be classified leads to prediction of low performance then it is clear that either the feature extraction process should be changed or that poor results must be

accepted. A method by which the P_{cc} could be predicted was proposed. The method relied on the Bhattacharyya bound on the error probability and as a result made several assumptions about the nature of the underlying data. Despite this, the approach was regarded as worth perusing since radar ATR performance prediction is a highly novel area of research and any success would make a welcome extension to existing knowledge.

Chapter 4

Micro-Doppler Signatures

Many radar targets are composed of structures that exhibit their own limited motion in addition to the target's general motion. For example, the tracks of a tank rotate as the vehicle moves forward, or the limbs of a person swing as they walk. Since all parts of the target scatter the radar signal the moving components will induce their own Doppler shifts alongside that result from the gross target motion. These extra shifts create the Micro-Doppler Signature (μ-DS) discussed in this chapter.

The μ-DS is introduced by looking at the different parts of a target that may generate a signature. Target motion when manoeuvring is considered and its unsuitability for classification identified. The idea of micro-motions of the target structure is presented and it is seen how this can lead to a unique μ-DS for a class of targets. The underlying theory is investigated through the use of a simple target model. The outputs of the model are used to demonstrate how the μ-DS may be observed and to provide insight into the difficulties a μ-DS classifier will face. With the theory established the Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) data that is used extensively in this research is described, and the μ-DS it holds investigated.

4.1 Micro-Doppler Signature Theory

The Micro-Doppler Signature (μ-DS) is the intricate frequency modulation that is imparted on the backscatter signal by the moving components of a radar target (Chen et al., 2006). Section 2.3 discussed how the radar echo was subject to a phase shift given by

$$\phi = \frac{4\pi R}{\lambda_{\mathrm{c}}} = \frac{4\pi R f_{\mathrm{c}}}{c} \tag{4.1}$$

where R is the target range; λ_{c} the wavelength of the carrier signal; f_{c} the frequency of the carrier signal; and c the speed of propagation. If the range, R, varies with time then so does the phase shift, ϕ. Frequency is the time derivative of phase so the changing phase results in a frequency shift (Stimpson,

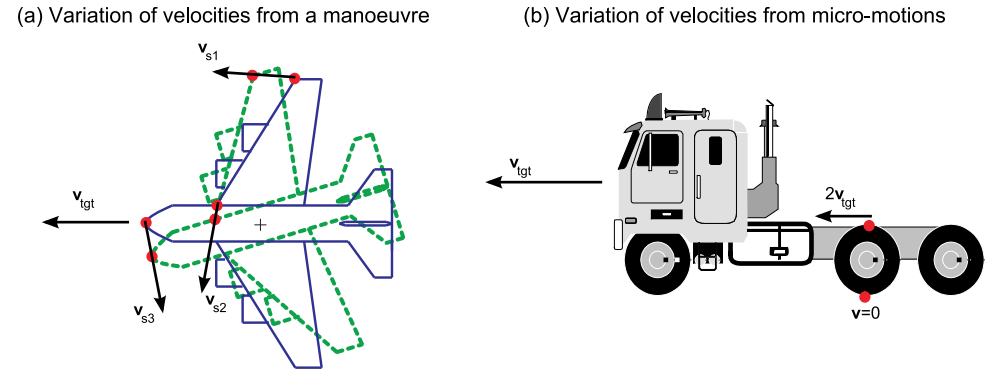


Figure 4.1: Two different mechanisms for creating a μ-DS: (a) as a result of the target manoeuvre and (b) as a result of micro-motions of target components.

1998). Generally the variation of R is considered to be a result of the target's radial velocity, v_{r}, and the frequency shift is the Doppler shift.

A target may be considered a collection of point scatterers and although it may seem counter intuitive there is no requirement that the velocities of the scatterers, $\mathbf{v}_{\mathrm{scat}}$, be the same as the target velocity $\mathbf{v}_{\mathrm{tgt}}$. The difference between the velocities may be attributed to one of two mechanisms:

1. The target may be rotating, such as when the rudder of an aeroplane is used to adjust its yaw, and scatterers at different distances from the centre of rotation will have different velocities, see Fig. 4.1 part (a).
2. The target may be built from components that can have their own, limited, micro-motions such as propellers, wheels or vibrating chassis parts. Fig. 4.1 part (b) shows the different velocities of motor vehicle: the top of each wheel is moving with twice the main target velocity while the wheel bottoms are stationary.

Each scatterer will cause its own Doppler shift of the backscatter signal and there will then be several frequency components that will interfere to generate a unique Doppler signature for the target.

The former mechanism results in a μ-DS that is not suitable for use in classification as it is dependent on the manoeuvre being performed rather than the target class. (Indeed, some classification approaches consider this movement as noise that must be removed (Rihaczedk and Hershkowitz, 2000).) The latter mechanism, however, results in a μ-DS that depends on the relationship between the point scatterers of the target and is therefore unique to the target. This makes it a strong candidate for being a characteristic signal that could be used in radar Automatic Target Recognition (ATR).

The spectral components of a μ-DS vary greatly in amplitude, especially compared to the amplitude of the Main Doppler Line (MDL), requiring the radar to have a very low noise floor and high dynamic range for them to be detected (Marple, 2001). In the data provided by Thales Ltd. at Crawley for use in this study, that will be discussed in detail in section 4.4, it is possible to see a difference of ≈ 40 dB between the components of the μ-DS. Despite the need for high dynamic range systems a form of μ-

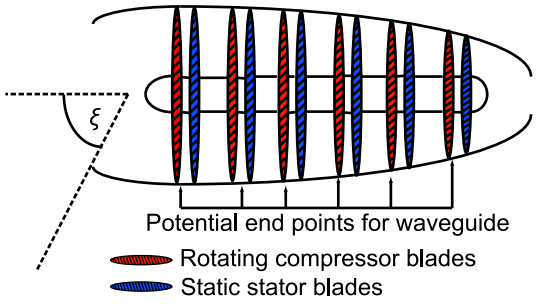


Figure 4.2: The jet engine considered as a JEM generating structure.

DS—Jet Engine Modulation (JEM)—has been discussed for many years in the radar literature (Gardner, 1961; Hynes and Gardner, 1967).

JEM is closely related to μ-DS and results from the interaction of the radar signal and the jet engine turbine. The turbine comprises of several layers of fans that compress the air as it is drawn into the engine. Each fan is a collection of rotating blades and as the signal is reflected from them they will impart their own Doppler shift, in the same manner as the point scatterers in the description above. There is a strong connection between JEM and the μ-DS of a propeller that is described in section 5.1 since each fan may considered as a propeller. However, there are further considerations in the JEM case as there are multiple fans and they are mounted in the engine housing.

The housing is able to shield the blades from the view of the radar as they are generally recessed within the engine (Piazza, 1999). Although this is often cited as simply preventing the radar from seeing the blades once the angle between the blades and the LOS is beyond some limit the effect is actually more subtle. Fig. 4.2 shows maximum angle at which the blades may be observed as ξ; if the angle is imagined increasing from 0° to ξ it is clear that the amount of blade observed will decrease due to occlusion by the housing. Since the obscured part of the blades is on one side of the engine only the resulting JEM will be asymmetrical in frequency.

The engine housing also acts as a waveguide with one closed end. The simplest interpretation is that the housing results in a delay, and hence phase difference, in the backscatter signal (Rihaczedk and Hershkowitz, 2000). The signal enters the housing/waveguide and travels along it before being reflected at the closed end. The extra distance travelled in this process results in a phase difference compared with signal reflected from the front of the engine. The closed end of the turbine is formed by the components within the engine e.g. the compressor blades and stator blades as indicated in Fig. 4.2, causing the waveguide length to vary and multiple phase differences. The motions of the internal components also cause the effective waveguide aperture to vary. The aperture at the front of the jet engine constantly varies as the compressor blades rotate before the stationary stator blades, the first pair of blades in Fig. 4.2. When the blades align with each other the aperture is largest and when they are completely unaligned the aperture is potentially closed. Models for a closed-ended, aperture-varying waveguide

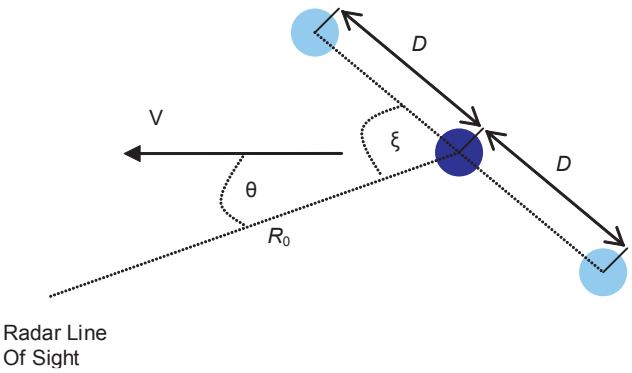


Figure 4.3: Configuration of a simple target that will generate a μ-DS.

have shown a jet engine housing to support many modes of the backscatter signal each with different phases (Tong et al., 1998).

Clearly the JEM effect is very complex and depends on a number of phenomena in addition to the Doppler effect. The use of JEM as an input for airborne radar ATR is an established technology (Skolnik, 1980; Bell and Grubbs, 1993; Martin and Mulgrew, 1990, 1992). The research presented in this thesis is concerned with ground targets that don't use jet engines. As such the μ-DSs that are considered result from target micro-motions: these are the focus of the next section.

4.2 Signature Of A Vibrating Scatterer

A simple target that will generate a μ-DS consists of two point scatterers: one of which has a micro-motion in addition to the general motion of the target. Such a target is shown in Fig. 4.3. Initially, at $t = 0\,\mathrm{s}$, both scatterers are a distance R_0 from the radar and are co-located (indicated by the dark blue circle). Both have a velocity, v, at an angle θ to the radar Line Of Sight (LOS). The projection of the velocity on the LOS is the target radial velocity, $v_\mathrm{r} = v\cos(\theta)$. For the first scatter this is a complete description of its motion, but for the second the micro-motion must also be considered. In addition to its linear velocity the scatterer is also vibrating (the limits of the vibration are indicated in Fig. 4.3 with light blue circles). The magnitude of the vibration is D, the vibration rate is ω and its initial phase is ψ. The expression $D\sin(\omega t + \psi)$ therefore provides a mathematical description of the micro-motion. If the vibration is at an angle ξ to the radar LOS, then its radial projection will be $\cos(\xi)\, D\sin(\omega t + \psi)$. Using the above an expression for the radial range as a function of time for each particle, $R_1(t)$ and $R_2(t)$ respectively, may be written:

$$R_1(t) = R_0 + v_\mathrm{r} t \tag{4.2}$$

$$R_2(t) = R_0 + v_\mathrm{r} t + \cos(\xi)\, D\sin(\omega t + \psi) \tag{4.3}$$

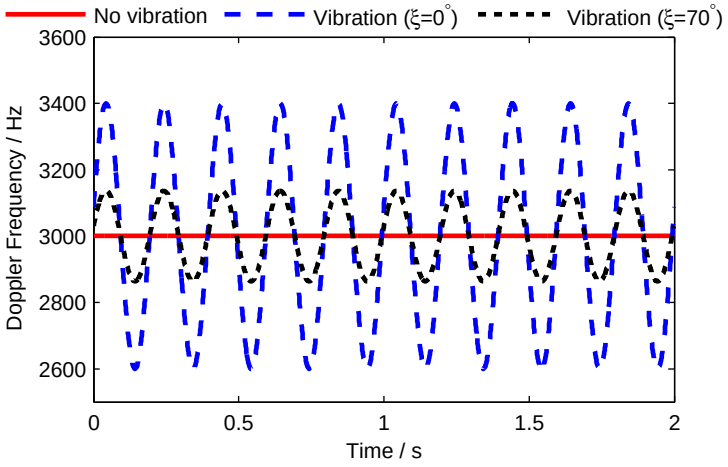


Figure 4.4: Variation of the Doppler frequency of a vibrating scatter.

Substituting (4.2) and (4.3) into (4.1) gives the expressions for the phase of the backscatter signal from each scatterer:

$$\begin{aligned} \phi_1(t) &= \frac{4\pi R_1(t)}{\lambda_c} \\ &= \frac{4\pi}{\lambda_c}\left[R_0 + v_r t\right] \qquad (4.4) \\ \phi_2(t) &= \frac{4\pi R_2(t)}{\lambda_c} \\ &= \frac{4\pi}{\lambda_c}\left[R_0 + v_r t + \cos(\xi)\, D \sin(\omega t + \psi)\right] \qquad (4.5) \end{aligned}$$

Differentiating these with respect to time gives the Doppler frequencies:

$$\begin{aligned} \dot{\phi}_1(t) &= \frac{4\pi v_r}{\lambda_c} \\ &= \frac{2\pi 2 v_r}{\lambda_c} \qquad (4.6) \\ \dot{\phi}_2(t) &= \frac{4\pi}{\lambda_c}\left[v_r + D\omega\cos(\xi)\cos(\omega t + \psi)\right] \\ &= \frac{2\pi}{\lambda_c}\left[2v_r + 2D\omega\cos(\xi)\cos(\omega t + \psi)\right] \qquad (4.7) \end{aligned}$$

The factors of 2π in (4.6) and (4.7) arise because the frequency is being considered is angular rather than linear.

The Doppler frequency of the scatterer with a micro-motion is time varying. Fig. 4.4 shows the Doppler frequencies obtained using (4.6) and (4.7). The parameters values are: $R_0 = 10\text{km}$, $v_r = 300\text{ms}^{-1}$, $\lambda_c = 0.2\text{m}$, $D = 8\text{m}$, $\omega = 5\text{Hz}$, $\psi = 45°$ and $\xi = 0°$ or $70°$ as indicated in the legend. When there is no micro-motion the Doppler shift is constant (the solid line); when the scatter vibrates the Doppler shift oscillates (the dashed and dotted lines). The centre of the oscillation is the frequency

associated with the radial velocity of the target, the $\frac{2\pi}{\lambda_c}2v_r$ term in (4.7). The remaining term in (4.7), $\frac{2\pi}{\lambda_c}D\omega\cos(\xi)\cos(\omega t+\psi)$, provides information about the frequency oscillation. All terms other than $\cos(\omega t+\psi)$ are constant and act to scale the cosine amplitude. The frequency of the oscillation must therefore be w and its starting phase ψ. $D\omega\cos(\xi)$ is the projection of the micro-motion onto the radar line of sight: ω is a frequency measured in rads^{-1} that is equivalent to an angular velocity so multiplying it by $D\cos(\xi)$, the projection of length D on the radar line of sight, converts it into a linear velocity along the line of sight. Multiplying this velocity by $\frac{2\pi 2}{\lambda_c}$ will give the Doppler frequency due to the vibration that is scaled by $\cos(\omega t+\psi)$. The minimum and maximum frequency values will occur at the limits of the cosine function (± 1) leading to:

$$\dot{\phi}(t)_{\max}=\frac{2\pi}{\lambda_c}\left[2v_r+2D\omega\cos(\xi)\right] \tag{4.8}$$

$$\dot{\phi}(t)_{\min}=\frac{2\pi}{\lambda_c}\left[2v_r-2D\omega\cos(\xi)\right] \tag{4.9}$$

By increasing the number of point scatterers used, assigning each one a Radar Cross Section (RCS) and defining the parameters of the illuminating radar in more detail it becomes possible to use (4.6) and (4.7) as the basis for a simple μ-DS simulator. The radar range equation, (2.3) from section 2.2, can be used to estimate the power of the backscatter signal from each point scatterer and coherently summing the signals provides the received signal from the target as a whole. The limit of this simulation is the micro-motion that each scatterer may exhibit: they can either have a vibration or no micro-motion at all.

Table 4.1 defines the parameters of the point scatterer μ-DS simulation developed. The simulation operates in a two dimensional space with the radar at the origin. The radar is defined simply in terms of the carrier frequency, transmit power, antenna gain and system losses. The main target is defined by the range zero and velocity parameters that govern the starting point and gross velocity of the target. Typically the values of these parameters would be the same for each point scatterer, although there may be a small variation in the value of range zero to simulate target extent, or the overall target would break up. The direction and amplitude of the micro-motion for a particular scatterer are specified simultaneously using an $x\hat{\mathbf{i}}+y\hat{\mathbf{j}}$ vector and the rate of oscillation is set through a vibration rate parameter. The starting phase of the micro-motion is set per scatterer allowing all configurations of the target to be simulated. Also associated with each point scatterer is the RCS value. The simulator has been used to demonstrate the effect of starting phase of the micro-motion and angle of observation on the μ-DS.

Three simulations with two vibrating scatterers were conducted: the first two investigated changing the initial phase of the micro-motion and the third the changing the observation angle ξ. Tables 4.2 to 4.3 show the parameters for the scatterers used in the simulations. In simulations one and two are described in Table 4.2 and Table 4.3 respectively, the only difference between them is that the initial phase of the vibrating, second, scatter is 90° rather than 0°, in the second simulation. The parameters

Parameter Name	Unit	Description
		Radar Parameters
Carrier Frequency	Hz	The frequency of the radar carrier signal.
Power	dB	The transmission power of the radar (before antenna gain).
Gain	dB	The gain of the radar antenna.
Losses	dB	Power losses in the system.
		Target Parameters
Range zero	m	The initial range of the target entered as a point $(R_0^{(\mathrm{i})}\hat{\mathbf{i}}, R_0^{(\mathrm{j})}\hat{\mathbf{j}})$.
Velocity	ms^{-1}	The velocity of the target entered as an $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ vector.
Vibration	m	The magnitude and direction of the vibration specified as an $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ vector.
Vibration Rate	Hz	The rate of a vibration.
Initial Phase	deg	The starting phase of the vibration.
RCS	m^2	The radar cross section of the scatterer.
		Simulation Parameters
Propagation Speed	ms^{-1}	The signal propagation speed, typically 3×10^8.
Start Time	s	The time at which the simulation starts.
End Time	s	The time at which the simulation finishes.
Time Step	s	The size of the increment in which time develops; related to the sampling frequency as $f_\mathrm{s} = \frac{1}{\text{time step}}$.
Use Hanning Window	N/A	Allows the user to apply a Hanning window to the FFT.

Table 4.1: The parameters of the vibrating point scatterer μ-DS simulator.

		Scatterer	
	Unit	**1**	**2**
Range zero	m	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Velocity	ms^{-1}	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration	m	$0\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration Rate	Hz	0	5
Initial Phase	°	N/A	0
RCS	m^2	5	2

Table 4.2: Parameter values for the scatterers in the two-scatterer simulations travelling at 45° to the LOS.

		Scatterer	
	Unit	**1**	**2**
Range zero	m	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Velocity	ms^{-1}	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration	m	$0\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration Rate	Hz	0	5
Initial Phase	°	N/A	90
RCS	m^2	5	2

Table 4.3: Parameter values for the scatterers in the two-scatterer simulation with 90° phase difference.

		Scatterer	
	Unit	**1**	**2**
Range zero	m	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$70 \times 10^3\hat{\mathbf{i}} + 70\hat{\mathbf{j}}$
Velocity	ms^{-1}	$71\hat{\mathbf{i}} + 71\hat{\mathbf{j}}$	$71\hat{\mathbf{i}} + 71\hat{\mathbf{j}}$
Vibration	m	$0\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$0.71\hat{\mathbf{i}} + 0.71\hat{\mathbf{j}}$
Vibration Rate	Hz	0	5
Initial Phase	°	N/A	0
RCS	m^2	5	2

Table 4.4: Parameter values for the scatterers in the two-scatterer simulations with 0° phase difference.

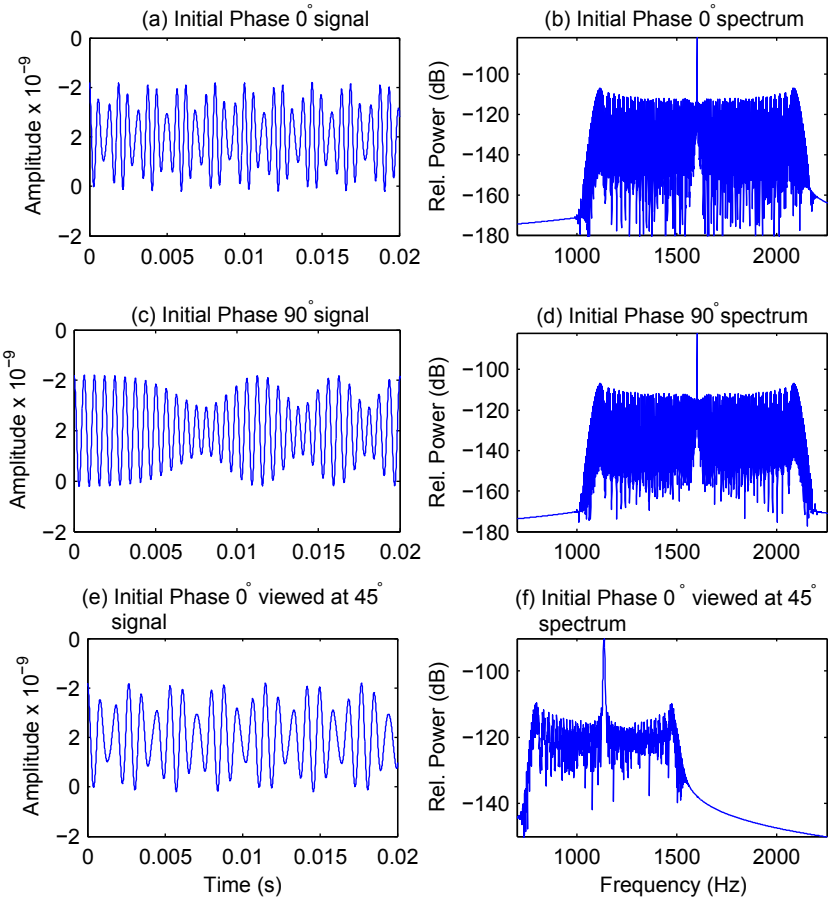


Figure 4.5: The real part of the signal and the power spectra for three two-scatterer simulations.

of the third simulation, described in Table 4.4, were the same as the first except for the direction of the velocity and vibration being rotated by 45° to simulate $\xi = 45°$. The parameters for the simulated radar were: carrier frequency $2.4\,\text{GHz}$, power $60\,\text{dB}$, gain $30\,\text{dB}$ and system losses $7\,\text{dB}$. The propagation speed was $3 \times 10^8\,\text{ms}^{-1}$, the simulation started at $0\,\text{s}$ and ended at $1\,\text{s}$ with a step of $10^{-5}\,\text{s}$. The output of the simulation is shown in Fig. 4.5 in which parts (a), (c) and (e) show a short sample of the real part of the received signal and parts (b), (d) and (f) show the corresponding power spectrum.

Between 0° and 90° phase cases there is no appreciable difference in the frequency domain, but substantial difference in the time domain. The backscatter signal of the second scatterer has a time varying frequency modulation resulting from the micro-motion. When this signal is summed with the constant frequency signal from the first scatterer they beat together causing the observable amplitude modulation of the received signal. The change in initial phase between the two cases means that the amplitude modulation is also different at a given moment in time although it varies with time in the same manner in each case. Because the power spectrum does not provide any time information, this difference is not visible and the two μ-DS appear identical. In the case where the target is viewed at 45° the received signal shows a similar amplitude modulation to 0° phase case, although the signal frequency

		Scatterer			
	Unit	**1**	**2**	**3**	**4**
Range zero	m	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Velocity	ms^{-1}	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration	m	$0\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration Rate	Hz	0	5	5	5
Initial Phase	°	N/A	0	120	240
RCS	m^2	5	1	1	1

Table 4.5: Parameter values for the four-scatterer simulations with even phase distribution.

		Scatterer			
	Unit	**1**	**2**	**3**	**4**
Range zero	m	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100 \times 10^3\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Velocity	ms^{-1}	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$100\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration	m	$0\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$	$1\hat{\mathbf{i}} + 0\hat{\mathbf{j}}$
Vibration Rate	Hz	0	5	5	5
Initial Phase	°	N/A	30	140	300
RCS	m^2	5	1	1	1

Table 4.6: Parameter values for the four-scatterer simulations with uneven phase distribution.

appears lower. This similarity and frequency variation is also seen in the power spectrum and is attributed to the view angle causing the radial component of the velocity to be less than in the head-on case. These results show the μ-DS to have a dependency on the particular observation with both the point in the micro-motion that the target is first detected and the aspect angle influencing the resulting signature.

Two further simulations using four vibrating scatterers were performed. In the first of these the initial phase of scatterers exhibiting micro-motions were evenly distributed between 0° and 360°, the parameters are detailed in Table 4.5. In the second case, detailed in Table 4.6, the distribution of scatterers phases was uneven. The results of these simulations are shown in Fig. 4.6. This time only the power spectra are shown with the full spectrum in parts (a) and (c) of Fig. 4.6 and a "zoomed in" view in parts (b) and (d). In both cases the bandwidth of the signature—the width of the power spectrum—is the same at approximately 1000 Hz. In the "zoomed in" views it is clear, however, that the fine structure is different. The signature is made up of a series of spectral lines that have a separation of 15 Hz when the initial phase is evenly distributed, see part (b) of the figure, and 5 Hz when the distribution is uneven, see part (d). For the even distribution the line spacing is:

$$\Delta f = N f_{\mathrm{v}} \tag{4.10}$$

where N is the number of scatterers with a micro-motion and $f_{\mathrm{v}} = \omega/2\pi$ is the linear frequency of the vibration. The result is consistent with the spacing of the lines in a propeller's spectrum (Schneider, 1987; Martin and Mulgrew, 1992; Mulgrew et al., 2003) where it is attributed to the backscatter signal from different rotor blades destructively interfering. By comparing the individual scatterers in this simulation

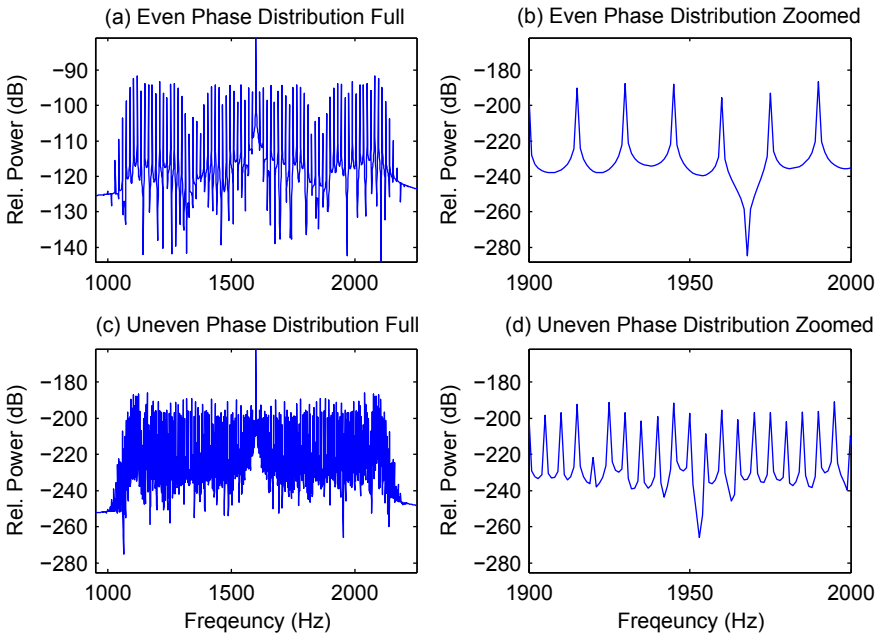


Figure 4.6: Power spectra for targets comprised of four point scatterers.

with the rotor blades it is straightforward to see why evenly distributed initial phases have a spacing given by (4.10).

4.3 Observing The Doppler Signature

The conventional approach to investigating the frequency spectrum of a discrete signal is to perform the Discrete Fourier Transform (DFT), usually achieved through application of the Fast Fourier Transform (FFT) algorithm (Gauss, 1866; Cooley and Tukey, 1965). This is the approach taken in section 2.3 to determine velocity of targets when using a pulse-Doppler radar. The question now is, "How suitable is the DFT when investigating the μ-DS?"

The DFT is defined as:

$$X[f] = \sum_{t=0}^{N-1} x[t] e^{-i2\pi f \frac{t}{N}} \quad \text{where} \quad \{f \in 0 \cup \mathbb{N} \,:\, f < N\} \tag{4.11}$$

where $X[f]$ is the discrete Fourier transform of $x[t]$, f is frequency, t is time and N is the number of samples of the signal. It is clear that $X[f]$ is a function of frequency only: the value associated with each $X[f]$ is the sum of the frequency components at f from *every* sample in $x[t]$ without any consideration given to when the frequency component was present. Each entry of $X[f]$ is effectively the sum, over t, of the projections of all the frequency components of $x[t]$ onto the phasor represented by $e^{-i2\pi ft/N}$. In the previous section it became clear that the μ-DS is a time varying phenomenon—the spectral components of the received signal vary with time. This was of particular significance in the two scatterer simulation when the notion of the backscatter signals "beating" against each other was discussed. Although the

analysis of the power spectrum was valuable it would have been better to be able to see variation of the spectrum with time.

Time-frequency analysis uses functions with two variables, one to represent time and the other frequency, the most common of which is the Short Time Fourier Transform (STFT) (Chen and Ling, 2002). By assuming that the signal under analysis is constant in frequency over short intervals it becomes acceptable to perform a Fourier transform on each interval without concern about the effect of frequency variation on the spectrum produced. The discrete STFT is defined as:

$$\mathrm{STFT}\,(\tau, f) = \sum_{t=0}^{N-1} x[t]w[t-\tau]e^{-i2\pi f\frac{t}{N}} \quad \text{where} \quad \begin{matrix} \{f \in 0 \cup \mathbb{N} : f < N\} \\ \{\tau \in 0 \cup \mathbb{N} : \tau < N - M\} \end{matrix} \tag{4.12}$$

where f and t are frequency and time as before, $x[t]$ is the signal of interest, $w[n]$ is the window function used to divide $x[t]$ into intervals, N the number of samples and M the length of $w[n]$. In (4.12) the window moves across the signal one sample at a time, however, in practical implementations the step size is often larger than a single sample. There is no requirement on the window to be a simple rectangle and it is common to apply weighting across the window. The window should ensure that the signal tapers smoothly to zero at each end to prevent smearing of the spectrum produced by a discontinuity between the first and last sample (Proakis and Manolakis, 1996). Typically the values of the STFT will be complex; a plot of the absolute values is call a *spectrogram* and provides an excellent mechanism for observing the μ-DS. Examples of using the STFT to analyse the μ-DS from the simulation of two scatterers from section 4.2, is shown in Fig. 4.7.

In Fig. 4.7 there are four different views of a μ-DS. The signal was for a target comprised of two point scatterers, of which one has a micro-motion—the target is described by column 1 and 2[a] of Table 4.2. The signal contains a linear frequency component at 1.6 kHz and components in the range $[1.0, 2.2]$ kHz that vary according to the $\cos(\cdot)$ term of (4.7). When the DFT of this signal is taken, shown in part (a) of Fig. 4.7 the stationary component at 1.6 kHz results in a clear peak in the power spectrum. Conversely, the non stationary frequency components result in a series of peaks throughout the frequency range $[1.0, 2.2]$ kHz giving the incorrect impression that the signal is comprised of a series of discrete, continuous tones. The STFT, using a rectangular window function of length 2048 samples, is shown in part (b). It is now possible to determine that the signal has a stationary frequency component as well as a time dependent one. Unfortunately the reduced length of the DFT performed at each time step has decreased the frequency resolution and the rectangular window has resulted in significant sidelobes appearing around the signature. Application of a Hanning window is able to reduce the level of the sidelobes in the STFT and such a transform is shown in part (c) of Fig. 4.7. The application of the windowing function has resulted in a further reduction in the frequency resolution and the individual

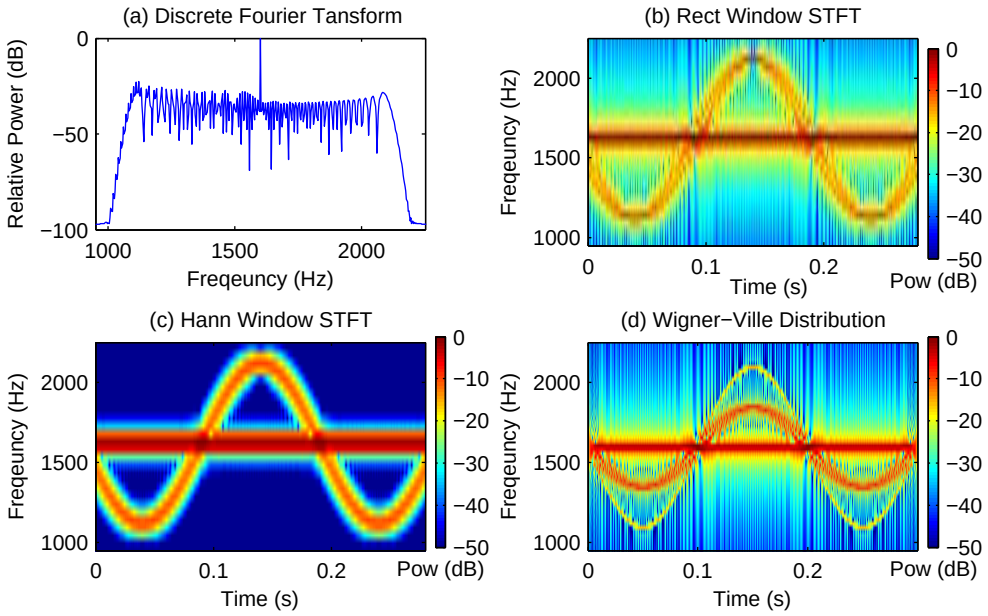


Figure 4.7: Comparison of viewing the same μ-DS like signal with (a) a DFT, (b) a rectangular windowed STFT, (c) a Hanning windowed STFT, and (d) the Wigner-Ville distribution (a high-resolution time-frequency representation).

peaks are much broader. Despite this loss in resolution the structure of the signal is very clear and there are limited appreciable sidelobes. Part (d) of the figure shows the output when a high resolution time frequency representation is used to process the signal.

The Wigner-Ville distribution is a quadratic time-frequency representation of a signal (Chen and Ling, 2002) that achieves superior resolution to the Fourier transform. The discrete version is defined as:

$$WVD[t,f] = \sum_{\tau=-\frac{N}{2}}^{\frac{N}{2}-1} x[t+\tau]x^*[t-\tau]e^{-i2\pi\tau f/N} \quad \text{where} \quad \{f \in 0 \cup \mathbb{N}; f < N\} \tag{4.13}$$

and is the Fourier transform of the time-dependent autocorrelation of $x[t]$. Fig. 4.7 part (d) shows the Wigner-Ville distribution of the signal defined above and it is clear that the frequency resolution is much finer than the STFT. Unfortunately when there is more than one frequency component present the Wigner-Ville distribution produces cross-terms and these too are seen in part (d) of the figure as the fine structure between the two genuine frequency components. These cross-terms arise because unlike the Fourier transform the sum of the Wigner-Ville distributions for two signals *is not* the same as the Wigner-Ville distribution for the sum of the two signals. It turns out that all of the super resolution time-frequency representations suffer from these cross-terms (Kootsookos et al., 1992). These cross-terms reduce the dynamic range of the time-frequency representations limiting their use for analysis of the μ-DS since it is not possible to determine which components are signature and which are cross-terms (Marple, 2001). Since the primary focus of the work presented in this thesis regards the investigation

of radar ATR, rather than advanced analysis of the μ-DS, it was decided that the STFT would be used to observer target signatures when required. With a method for observing the μ-DS decided upon it is possible to investigate some real signatures.

4.4 Analysis Of Thales MSTAR Data

Monostatic data recorded by a Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) radar operating at Ku-band was available. Three classes of ground target were measured: a wheeled vehicle, a tracked vehicle and personnel. Each target was recorded at several aspect angles ranging between head-on (0°) and tail-on (180°). The measurement interval was 30° for vehicles and 45° for personnel. Clutter profiles of the test range were also recorded. The recorder collected coherent data from a series of contiguous range bins. Each run produced several seconds of data, typically limited by the amount of time it took the target to traverse the range bins. A small amount of pre-processing was required to reconstruct a continuous signal rather than one spread across range bins.

A reconstructed signal, sampled at the radar's pulse repetition frequency (PRF), was produced by coherently summing the range bins. The summation effectively created one large range bin and overcame the need to track the target as it moved between bins. The reconstructed signal was then manually checked for the presence of interference or excess noise in the data channels by observing the signal's spectrogram. An example of each target at each aspect angle was chosen in which there were several seconds of data and a strong μ-DS. Once the signals have been reconstructed they are suitable for application of the DFT and STFT to observe the μ-DS.

Figures 4.8 to 4.13 show the μ-DS for three different classes of target: tracked vehicles, wheeled vehicles and personnel. Each of the figures is broken into three parts: part (a) shows the power spectrum for the entire reconstructed signal; part (b) the power spectrum for 1024 samples selected from the middle of the reconstructed signal; and part (c) shows the spectrogram. In all of the power spectra there is a strong ground response at zero on the frequency axis. Each target class is shown travelling head-on to the radar LOS and then again at an angle to the LOS. For the vehicles the angle is 60° and for the personnel 45°. The two vehicles have a velocity in the range 20 to 50 MPH (precise values were not logged during the trial) while the personnel target was running. The features of each target class will be considered in turn.

The tracked vehicle μ-DS shown in Figs 4.8 (head-on) and 4.9 (60°) has a distinctive signature. In the spectrogram, part (c), there is a strong Doppler response varying between 0.22 and 0.27 on the normalized frequency axis that corresponds to the main chassis of the vehicle. The frequency variation is a result of the speed changing during the trial. Surrounding the main line is a series of lines that constitute the target's μ-DS. It is assumed that these lines are reflections from the caterpillar tracks of the vehicle. Supporting this assumption is the presence of lines between 0 Hz and the chassis frequency: the part of

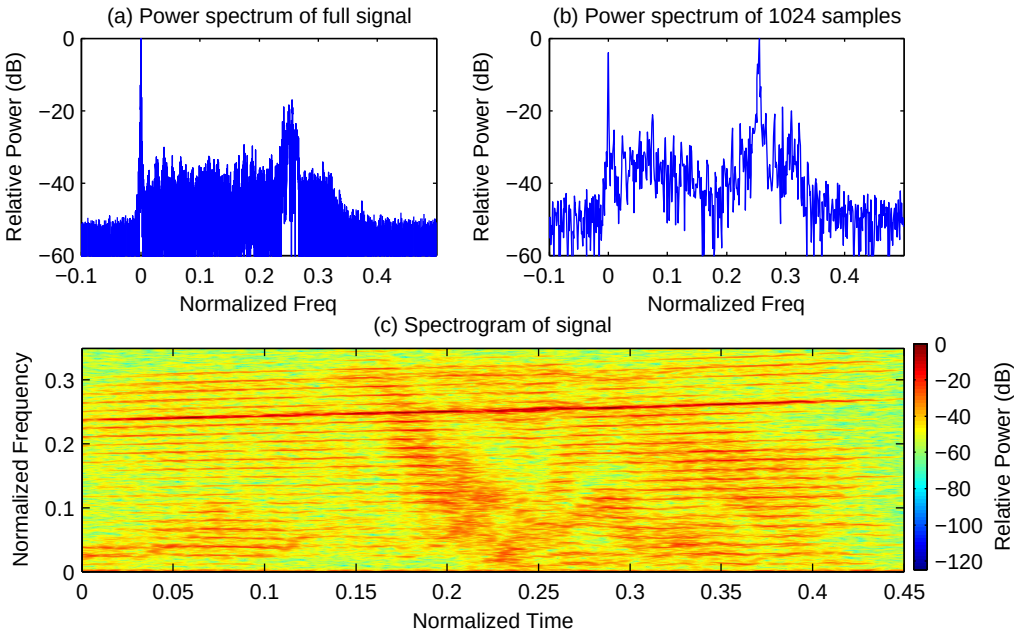


Figure 4.8: The μ-DS for a tracked vehicle travelling at 0° to the radar LOS.

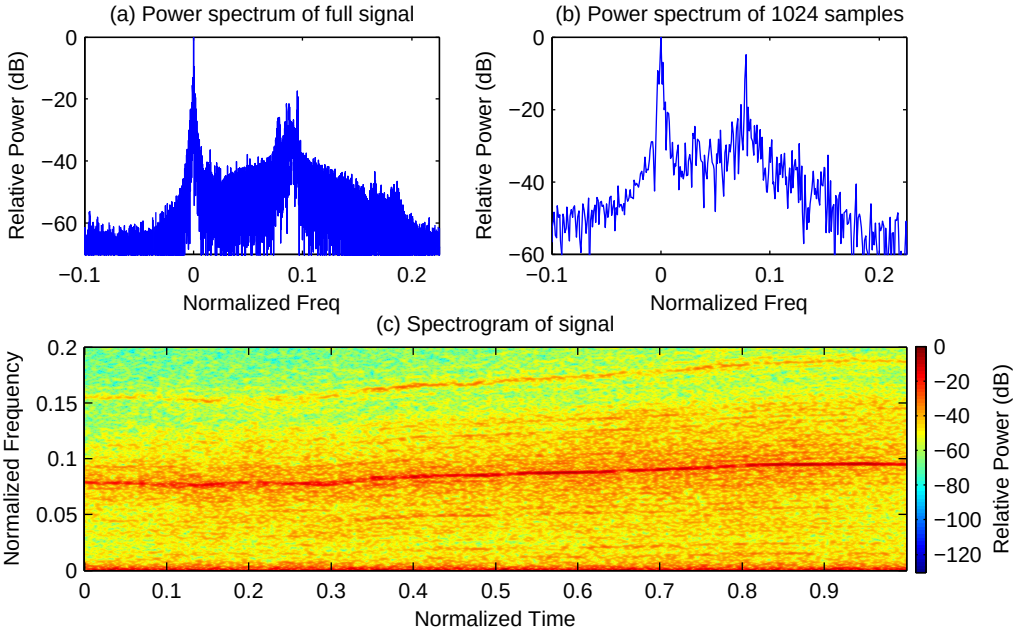


Figure 4.9: The μ-DS for a tracked vehicle travelling at 60° to the radar LOS.

the track that is in contact with the ground has a speed of 0 MPH and so doesn't cause a Doppler shift. The top part of the caterpillar track will be travelling with twice the speed of the main body so it was expected to observe frequencies up to twice that of the chassis in the μ-DS. However, the upper sideband of the signature appears to stop prematurely and this was attributed to the panniers of the vehicle covering the top part of the tracks and shielding them from the radar's transmission. It is possible that the number of spectral lines in the signature and their spacing provide information about the target's identity, as they do for the vibrating scatterer model in section 4.2 and for propeller blades (Schneider, 1987; Martin and Mulgrew, 1990; Bell and Grubbs, 1993), see section 5.1. Unfortunately, without detailed knowledge of the physical properties of the target it is unclear how to extract this information from the signature. The two DFTs support the observations of the spectrogram with both showing a strong spectral response at the chassis frequency with an asymmetrical μ-DS around it. The signature of part (b), the 1024 point DFT has better definition because the μ-DS was constant in frequency over the samples from which the spectrum was calculated. In section 4.3 it was discussed that the DFT is not suitable for processing time varying signals and variation in target speed cause a time variation in the frequency spectrum. The peaks of the full signal duration DFT "smear" in frequency as a result of this variation. When the target is turned to 60° from the radar LOS there are two principal changes to the μ-DS, both of which can be seen in the spectrogram of Fig. 4.9 part (c). The frequency of the MDL is reduced since the radial velocity of the target has fallen compared with the head-on case and the μ-DS bandwidth has reduced causing the distinctive lines to merge into one and other causing a broad spectral response with less obvious structure. A single distinct line can be seen at twice the chassis frequency that was attributed to the top of the caterpillar track. Now the tank is being illuminated from an angle the top of the track is no longer shielded. These observation angle changes in the μ-DS were most significant for the wheeled vehicle target.

Figures 4.10 and 4.11 show the μ-DS for the wheeled vehicle target. There is a clear distinction between the two figures: when the target is travelling head-on to the radar LOS the μ-DS bandwidth is so small as to be nearly non existent. In the head-on case only the MDL resulting from the vehicle chassis can be seen. (There are two small μ-DS components of the signal at ≈ 0.17 and ≈ 0.23 normalized frequency and between 0.1 and 0.2 normalized time but they are over 30 dB below the MDL.) In the 60° case, however, the μ-DS is visible as a series of lines surrounding the chassis response. The difference between the two views is a result of the wheels being shielded from the radar by the chassis. In Fig. 4.11 the μ-DS extends from 0 to twice the frequency of the MDL suggesting that it is being generated by the wheels of the target. When the target is travelling head-on to the radar the wheels are shielded by the wheel arches and mud flaps. At 60°, however, the radar can "see into" the wheel arches and there is a response from the wheels resulting in a μ-DS. In a manner similar to the tracked vehicle when illuminated from the side the signature is symmetrical about the chassis line suggesting even illumination

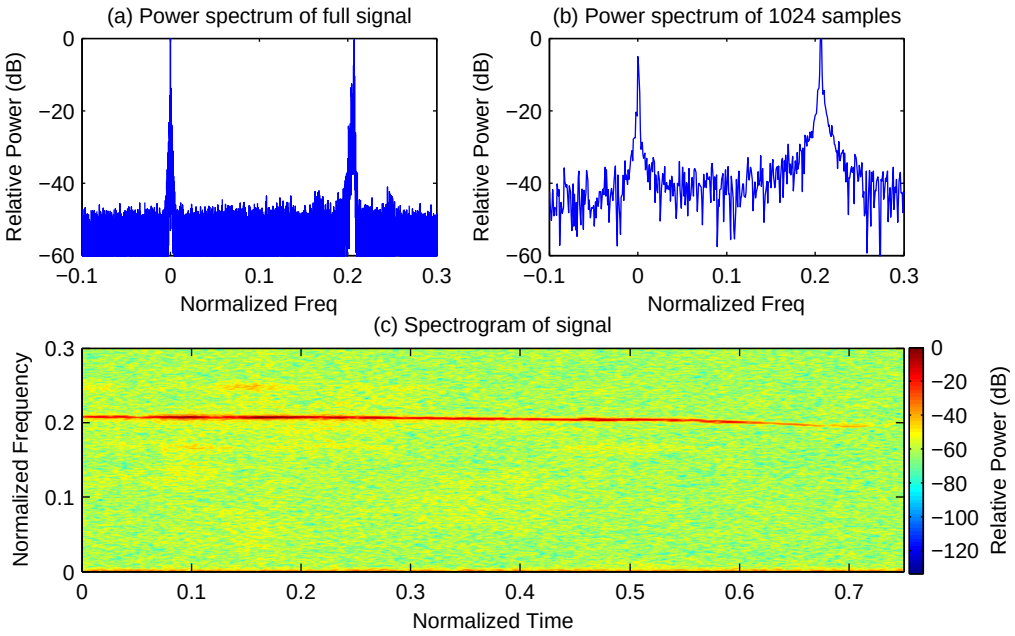


Figure 4.10: The μ-DS for a wheeled vehicle travelling at 0° to the radar LOS.

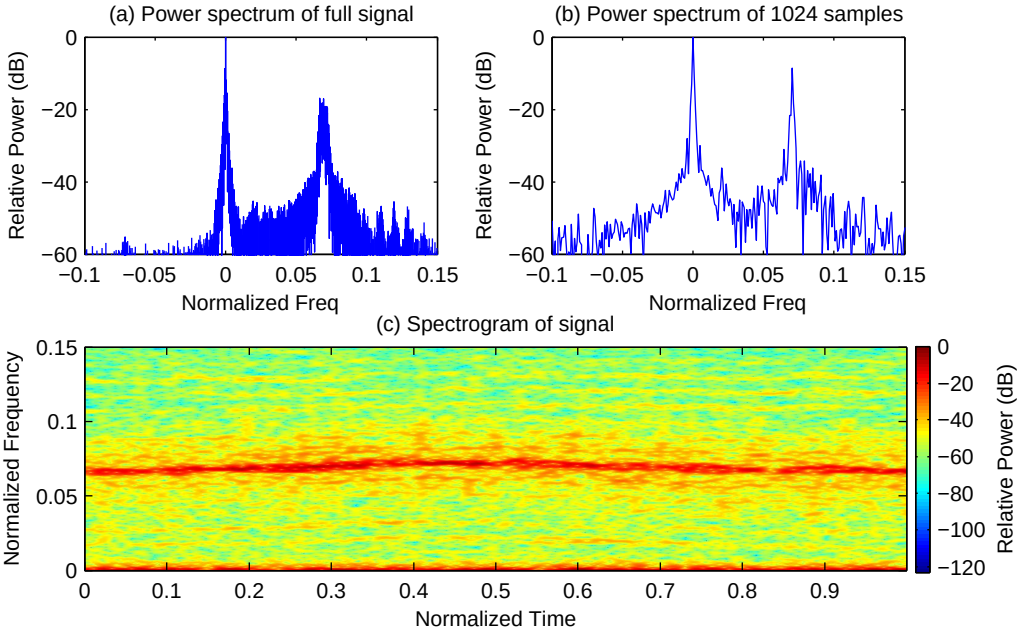


Figure 4.11: The μ-DS for a wheeled vehicle travelling at 60° to the radar LOS.

of the wheel.

The last target class is personnel and the result of two trials, one head-on and the other at 45° to the radar LOS, are shown in figures 4.12 and 4.13. When the target is head-on to the radar LOS there is a distinctive signature. Surrounding the MDL, that is a result of the target body motion, are oscillating frequency components that are comparable to the response for vibrating scatterers (see section 4.2). These components correspond to the forward and backward motion of the arms and legs of the running person. When the target is at an angle of 45° to the LOS the signature becomes less distinctive and the bandwidth is reduced. The only components of target motion that contribute to the signature are those along the LOS; therefore the range of motion detected by the radar from the target limbs is reduced and the signature bandwidth falls. With the constant variation in frequency that results from oscillating motion the DFT is particularly unsuited for analysing the μ-DS of personnel. In both the full signal and 1024 samples power spectra, part (a) and (b) of the figure, the spectral lines have been smeared into one and other and it difficult observe much more than the signature bandwidth. In contrast to this the spectrogram contains much information about the target: it is possible to estimate the *step frequency*, the rate with which the limbs swing, from the varying frequency components of the μ-DS. Step frequency is a critical parameter of human locomotion (Boulic, 1990) and can be used to identify individuals (van Dorp and Groen, 2003).

4.5 Summary

The basic theory behind the μ-DS was explained in this chapter, both mathematically and using real data examples. It was reported that the micro-motions of parts of a radar target resulted in the backscatter signal having a time-varying frequency modulation known as the μ-DS. The preliminary discussion explained how the JEM phenomena, that has been used for several years in target recognition, is a form of micro-Doppler. From here a mathematical description of the μ-DS was provided.

The μ-DS was explained through the use of a mathematical simulation of point scatterers that exhibited micro-motions. It was demonstrated that in addition to the basic parameters of the simulation, the starting conditions could have an effect on the observed signature. It was also shown that the nature of the signature could vary over time. Both of these properties were identified as challenges a radar ATR system would have to overcome. The time varying nature of the signature requires the use of time-frequency analysis techniques, rather than conventional Fourier analysis, in order to perform detailed investigation. However, it was also discussed that because of the low power of the micro-Doppler components in a signal use of high resolution time-frequency methods, such as the Wigner-Ville distribution, was not practical due to confusion between cross-terms and signature structure. With the basic theory developed real world examples were given.

The Thales MSTAR dataset, that is used extensively in this research, was investigated and its micro-

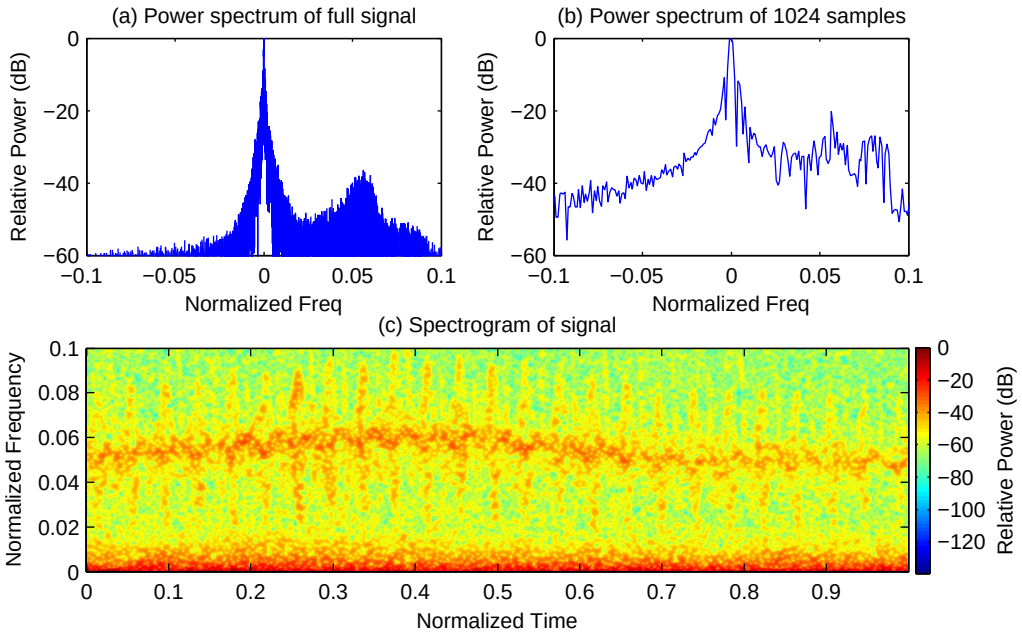


Figure 4.12: The μ-DS for a running person travelling at 0° to the radar LOS.

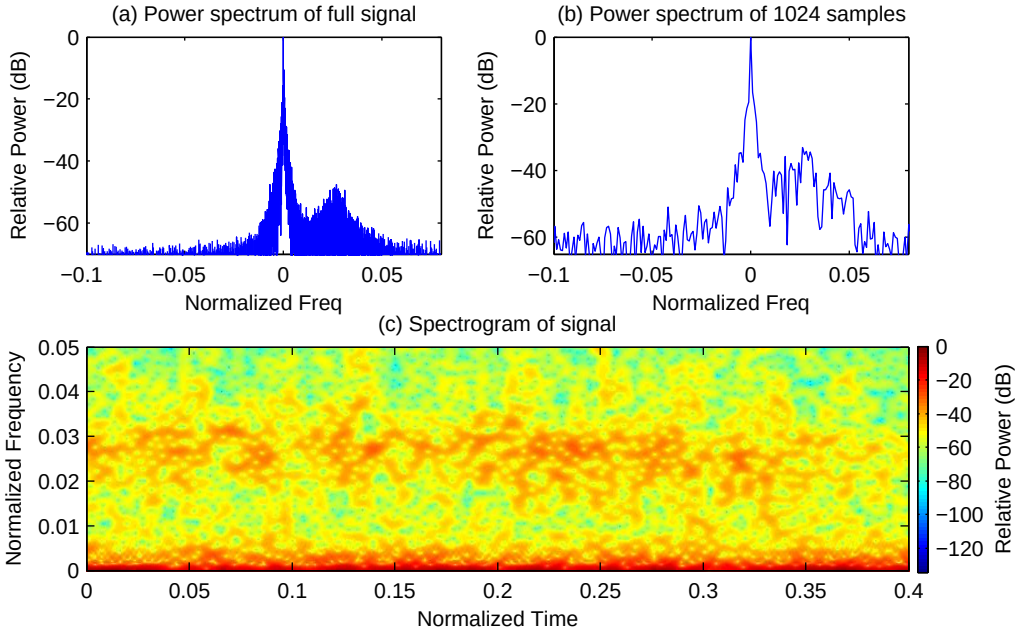


Figure 4.13: The μ-DS for a running person vehicle travelling at 45° to the radar LOS.

Doppler content assessed. Example spectrograms and power spectra for the three target classes is the dataset—tracked vehicles, wheeled vehicles and personnel—were presented. It was observed that the spectrograms were easily distinguishable to a human observer and this increased confidence that automatic recognition would be possible. Unfortunately, it was also observed that the signatures varied with target aspect angle. This variation was regarded as providing an additional challenge for potential radar ATR systems.

Chapter 5

Data Pre-Processing

The Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) data used in this study required pre-processing before it could be used in classification tests. This required the development of a number of new processing techniques. These are described in this chapter and can be divided into the following topics:

1. *The synthesizing of two additional target classes for use as "unknown" targets.*
2. *Division of the long duration reconstructed signals into short frames with durations comparable to a scanning radar dwell time.*
3. *Selection of data frames for use as either reference or test data.*
4. *Processing of the data to account for the classifier feature extraction stage.*

5.1 Synthesizing "Unknown" Targets

In section 3.2 it was identified that if a classifier's "unknown" input rejection capability is to be tested it will be necessary for there to be some target classes in the test dataset that are not in the reference dataset. The Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) data, that was introduced in section 4.4, provides examples of the Micro-Doppler Signature (μ-DS) for three classes of targets. Holding back one of the targets to use as an unknown was considered, but it was felt that reducing the classification to discrimination between two classes trivialized the problem. In addition to trials with the target present, the MSTAR data also contained several clutter profiles in which the radar was operated without a target. The "unknown" target examples were formed by combining two of these clutter profiles with data for simulated targets.

Two additional targets were synthesized. The first artificial signal comprised I and Q samples independently selected from a uniform distribution, this target will be referred to as the *random target*. The second was a simulation of the return from a propeller generated using the simulation described below; this target will be referred to as the *propeller target*. Prior to addition to the clutter profiles both signals were scaled. Firstly, so they had peak amplitudes similar to the tracked vehicle examples

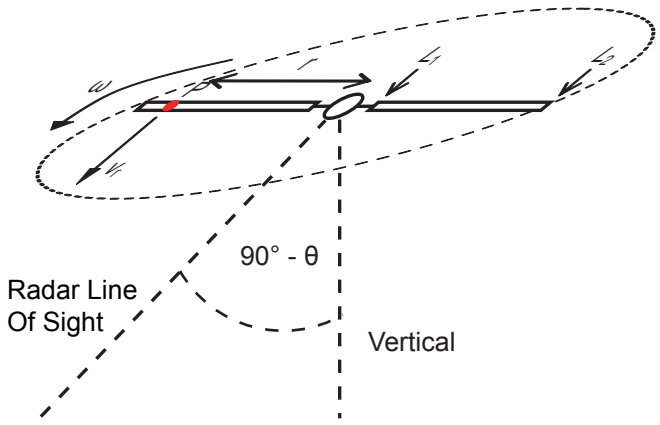


Figure 5.1: The geometry of a radar illuminating a propeller.

in the provided dataset, and secondly, so that each sample of the signals was stored using the same number of bits as the Analogue to Digital Converter (ADC) of the Thales MSTAR data recorder. Once the synthesized signals had been produced they were saved in the format of the original data so they could be subject to the same processing. Although the synthesized signals did not match any real world target but were still considered suitable for use since the signal values could be output by a radar. The random target was expected to be very different to the real data, and so easily detected as "unknown" by the classifiers. The propeller target had a micro-Doppler like quality to it and was expected to be more difficult to reject.

The basic structure of a propeller is a hub with blades that extend from it. The effect of the hub was ignored and the propeller was considered as a collection of blades that start a distance L_1 from the centre of rotation and end at a distance L_2, the length of a blade was given by $L_2 - L_1 = L$. The number of blades on the propeller was N. While there may be an odd or even number of blades, it was assumed that the blades are evenly spaced. The angular velocity of a blade was ω and is measured in rad.s^{-1}. Finally, the blade was modelled as a single rigid, homogeneous, linear antenna (Schneider, 1987) so the complex aerofoil shape and twist to the blade was ignored. As with the vibrating scatterer theory from section 4.2 the μ-DS depended in part on the radar-target geometry.

The geometry of a propeller being illuminated by a radar is shown in Fig. 5.1. The propeller rotates in a plane that is at an angle θ to the radar Line Of Sight (LOS). Each point P along a blade, that is distance r from the centre of rotation, has a linear velocity $v_{\mathrm{r}}^{(\mathrm{P})} = r\omega$ that results in a Doppler shift of any radar signal incident upon it. By integrating the Doppler shifts along the blade length and summing for each blade the expression for the radar echo signal is found to be (Schneider, 1987; Martin and

Mulgrew, 1990):

$$\begin{aligned} u(t) &= \sum_{n=0}^{N-1} A_{\mathrm{r}} (L_2 - L_1) e^{i\left(\omega_{\mathrm{c}} t - \frac{2\omega_{\mathrm{c}}}{c}\left[R + v_{\mathrm{r}} t + \frac{L_1+L_2}{2}\cos(\theta)\sin\left(\omega t + \frac{2\pi n}{N}\right)\right]\right)} \\ &\quad \times \mathrm{sinc}\left(\frac{2\omega_{\mathrm{c}}}{c}\left[\frac{L_2 - L_1}{2}\right]\cos(\theta)\sin\left(\omega t + \frac{2\pi n}{N}\right)\right) \end{aligned} \tag{5.1}$$

where A_{r} is a scale factor allowing for the blade Radar Cross Section (RCS); t is time; ω_{c} is the angular carrier frequency ($\omega_{\mathrm{c}} = 2\pi f_{\mathrm{c}}$); R is the radial range of the centre of rotation; v_{r} is the radial component of the linear velocity of the centre of rotation; and all other variable are as described above. The Fourier transform of (5.1) is:

$$U(f) = \sum_{k=-N_1}^{N_1} c_{Nk}\delta\left(f - f_{\mathrm{c}} - f_{\mathrm{D}} - Nkf_{\omega}\right) \tag{5.2}$$

where

$$\begin{aligned} c_{Nk} &= \sum_{l=0}^{\infty} \frac{2(-1)^{Nk\mathrm{u}(0)} A_{\mathrm{r}} N}{\frac{4\pi}{\lambda_{\mathrm{c}}}\cos(\theta)} \\ &\quad \times \left(\mathrm{J}_{|Nk|+2l+1}\left(\frac{4\pi}{\lambda_{\mathrm{c}}} L_2 \cos(\theta)\right) - \mathrm{J}_{|Nk|+2l+1}\left(\frac{4\pi}{\lambda_{\mathrm{c}}} L_1 \cos(\theta)\right)\right) e^{-i\frac{4\pi R}{\lambda_{\mathrm{c}}}} \end{aligned}$$

and $f_{\mathrm{c}} = \frac{\omega_{\mathrm{c}}}{2\pi}$ is the carrier frequency in Hz; $\lambda_{\mathrm{c}} = \frac{c}{f_{\mathrm{c}}}$ is the carrier wavelength; $f_{\mathrm{D}} = \frac{-2v_{\mathrm{r}}}{\lambda_{\mathrm{c}}}$ is the Doppler frequency; $f_{\omega} = \frac{\omega}{2\pi}$ is the frequency of rotation in Hz; $\mathrm{J}_k(\cdot)$ is a kth order Bessel function of the fist kind; $u(x)$ is the unit step function; $\delta(x)$ is the Dirac delta function; and N_1 is the highest significant sideband.

The model described in (5.1) and (5.2) is well understood (Schneider, 1987; Martin and Mulgrew, 1990; Pellegrini et al., 1995) and the following properties apply to it:

$$\Delta f = N f_{\omega} \tag{5.3}$$

Assuming $L_1 = 0$

$$N' = \frac{8\pi L_2 \cos(\theta)}{N\lambda_{\mathrm{c}}} \tag{5.4}$$

$$B = \frac{8\pi f_{\mathrm{r}} L_2 \cos(\theta)}{\lambda_{\mathrm{c}}} \tag{5.5}$$

Assuming $L_1 \neq 0$

$$N' = \frac{8\pi (L_2 - L_1) \cos(\theta)}{N\lambda_{\mathrm{c}}} \tag{5.6}$$

$$B = \frac{8\pi f_{\mathrm{r}} (L_2 - L_1) \cos(\theta)}{\lambda_{\mathrm{c}}} \tag{5.7}$$

where Δf is the spacing of the spectral lines; N' the number of significant sidebands, that may be used to select N_1 in (5.2); and B is the bandwidth of the signal. The factor of N in the divisor of

Parameter	Unit	Propeller Target
Start time	s	0
End time	s	10
Time step	s	Secret*
Propagation speed	ms^{-1}	3×10^8
Carrier Wavelength	m	Ku-Band*
N° Blades	N/A	2
Blade Start Length	m	0.15
Blade End Length	m	0.9652
Rotor Speed	RPM	2300 RPM
Blade Pitch	°	90
Body RCS to prop RCS ratio	dB	$-\infty$
Angle to LOS	°	0
Target Velocity	ms^{-1}	66.7

*The precise values for these parameters are subject to a non-disclosure agreement due to commercial sensitivity.

Table 5.1: The parameters used to simulate the μ-DS of a propeller structure.

equations 5.4 and 5.6 is a result of harmonic phase cancellation. Equation (5.5) and (5.7) are obtained by multiplying the relevant N' by Δf. Equation (5.2) shows that the spectrum is a discrete series of spectral lines, rather than a continuous frequency distribution. Essentially power spectrum is a collection of Dirac delta functions, spaced at intervals of Nf_ω, with amplitude modulation provided by the various c_{Nk} coefficients. The notion of the backscatter signal from a propeller being composed of discrete frequencies is supported by an alternative work that reaches the same conclusion from a more general derivation (Bell and Grubbs, 1993).

For the propeller target the variables of (5.1) were selected based on a Piper Archer II light aircraft, and the radar parameters were selected to match the Thales MSTAR, the exact values are listed in Table 5.1. Unfortunately, in Table 5.1 the radar and timing parameters are vague or marked as "secret" because they are deemed to be of a commercially sensitive nature. Different views of the propeller signal are shown in Fig. 5.2. The axis of the graphs have been normalized, because of commercial sensitivities. Part (a) of the figure shows the time domain signal to exhibit the blade flash phenomenon but also an intricate amplitude modulation. The modulation is a result of the radar Pulse Repetition Frequency (PRF), that acts as the sampling frequency, not being high enough to sample the Doppler shift resulting from the blade tips. As a result there is aliasing and the aliased frequencies "beat" against the lower frequencies from the parts of the propeller close to the hub causing modulation of the cumulative signal. This effect is also seen in parts (b) and (c) of Fig. 5.2 and results in the signal's power spectrum being a densely populated series of delta functions. With the transmit parameters of the Thales MSTAR the bandwidth of the signal, calculated using (5.7), is of the order $50\,\text{kHz}$ and results in maximum and minimum frequencies of interest with magnitudes in excess of three times the Nyquist frequency. Through aliasing and frequency fold over effects it becomes difficult to distinguish any meaning from the power spectrum or spectrogram since multiple Nyquist regions have be projected on top of each other. Fig. 5.3 shows the

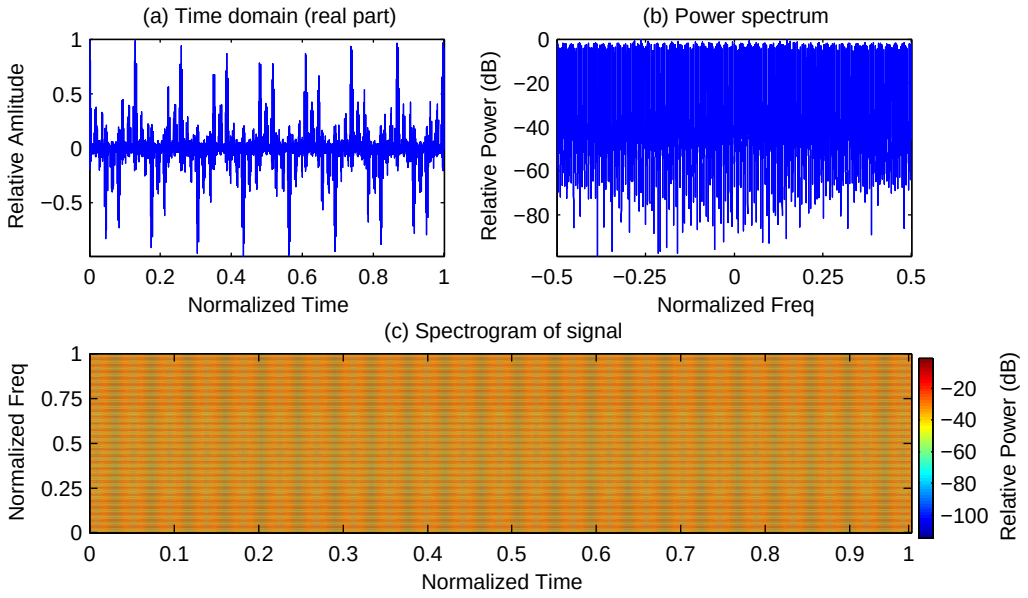


Figure 5.2: The output from the Piper Archer II propeller simulation. Part (a) shows the baseband signal; part (b) its Fourier transform; and part (c) the time-frequency representation.

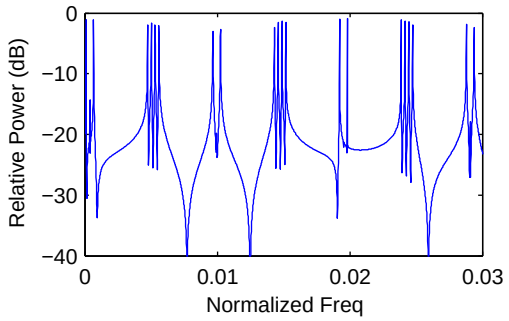


Figure 5.3: Power spectrum of the simulated propeller signal over a limited frequency range.

power spectrum of Fig. 5.2 part (b) over a limited frequency range. It is clear in this "zoomed in" view that the spectrum is a series of discrete spectral lines. It is observed that the rather than be at even spacing, as theory predicted, the lines a bunched together—a result of the fold over and aliasing difficulties described above.

With the two "unknown" targets successfully synthesized the Thales MSTAR dataset comprised of five classes: three real ones, from which both reference and test dataset would be taken; and two simulated, that would only provide examples for the test dataset. The next preparation stage was intended to increase the realism of the data. Until now reconstructed signals that are up to ten seconds long have been used to produce the various plots and figures presented. Data from an operational radar would not have such a long duration, so it was decided to divide the signals into short frames.

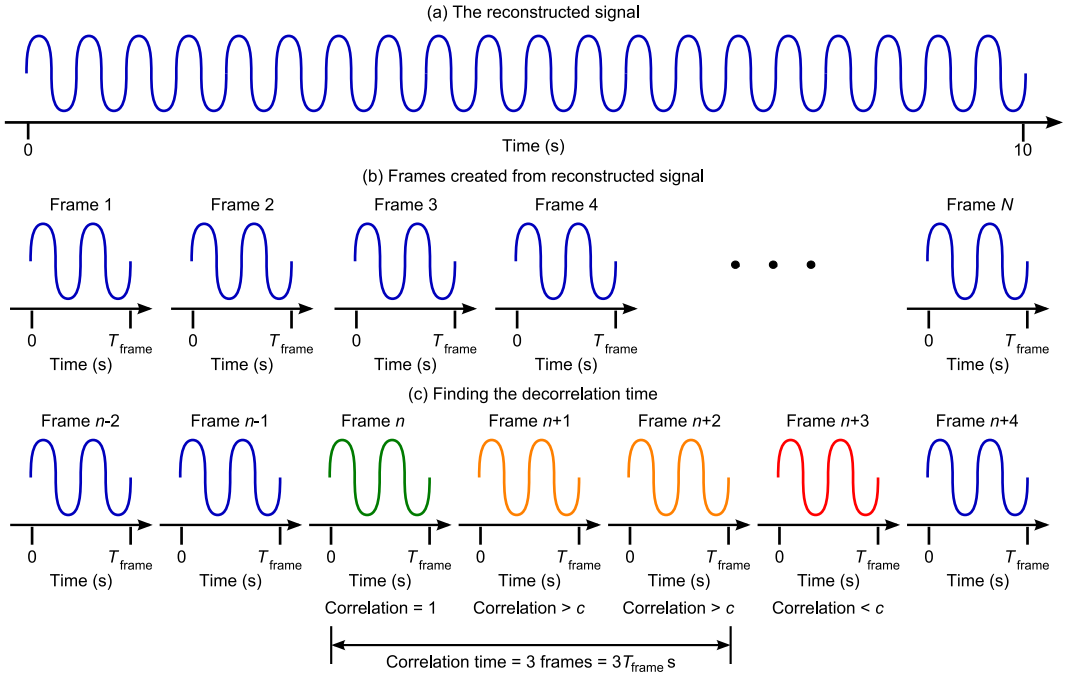


Figure 5.4: Dividing the reconstructed signal into realistic frames and measuring the correlation time.

5.2 Creating Realistic Duration Data

When a radar is performing a volume scan it will not illuminate any target for more than a few milliseconds (Stove and Sykes, 2003). It would be ideal if the Automatic Target Recognition (ATR) system could perform its identification within the short time frame to prevent interruption to scanning. The reconstructed signals that were introduced in section 4.4 are up to 10 s in duration and so do not reflect a realistic signal for a scanning radar. To increase realism these long signals were divided into short frames with durations more representative of the output of an operational radar.

A set of short frames was created from each of the reconstructed signals; the process for creating these frames is shown in Fig. 5.4. The reconstructed signal is shown in part (a) and will be considered as being comprised of P samples. The duration of the signal is a maximum of ≈ 10 s, far longer than a radar dwell time, so it is cut into short frames, as shown in part (b). The frames are contiguous in time, and if each one is Q samples long then number of frames created, N, is:

$$N = \text{floor}\left(\frac{P}{Q}\right), \tag{5.8}$$

where floor is a function that rounds to the nearest integer towards $-\infty$. If there is a short frame left at the end with less than Q samples in it is discarded. Clearly the duration of each frame will be Q divided by the radar PRF and in the study frame durations of 8 ms, 16 ms, 32 ms, 48 ms and 64 ms were used. Once the reconstructed signal has been cut into frames its correlation time can be measured, this is shown in part (c) of Fig. 5.4.

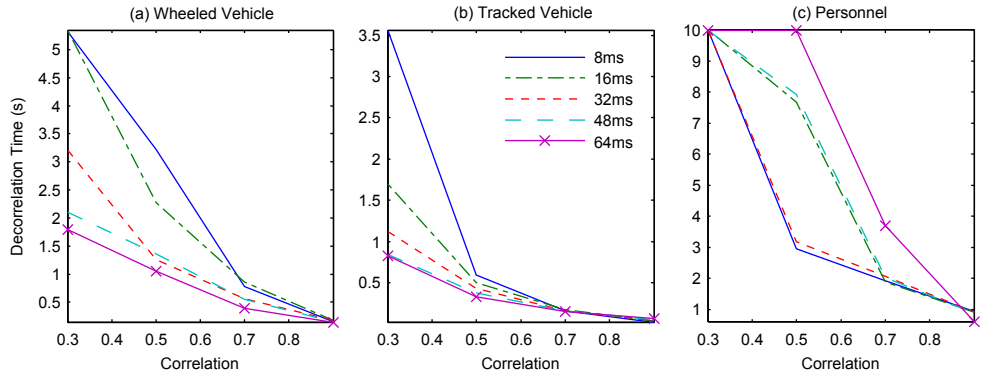


Figure 5.5: Correlation times for wheeled vehicle, tracked vehicle and personnel targets.

The signal correlation time is a measure of the duration for which the signal remains correlated with itself: it is the lag in the autocorrelation function before the correlation falls below a pre-determined level, c. Since the frames created above are contiguous it can be calculated from them as shown in part (c) of Fig. 5.4. The frames are arranged in order, one is selected and then correlated with its neighbours until the peak value of the correlation between them drops below c. The decorrelation distance is then the number of frames the signal remained correlated for and the correlation time is the number of frames multiplied by the frame duration, T_{frame}. In the figure the time is shown as being calculated by correlating the chosen frame with those that come after it in time. In practice a measure can also be made by correlating it with those that came before also. The final correlation time for the signal is the average of all the forward and backward correlation times for each frame. The measured correlation times for the three targets is show in Fig. 5.5.

In this research the point, c, at which a signal was regarded to have become decorrelated was considered as 0.3, 0.5, 0.7 and 0.9. The variation in the correlation time for each of the three targets, at each of the five frame durations used, is shown in Fig. 5.5. The correlation time for both vehicles is generally shorter than the time for the personnel. This is attributed to the limbs of the person moving much more slowly than the wheels or tracks of the vehicles. The slower motion will cause the μ-DS to change less rapidly and so remain correlated longer. For the personnel case the 64 ms frame having a correlation time of 10 s for $c = 0.3$ and 0.5 suggests that the correlation between the frames never fell to c. The impact of frame duration does not appear to have a pattern for personnel while for vehicles it is clear that longer frames have shorter correlation times. In all cases, as would be expected, when the value of c is highest, 0.9, and the frames must be very similar to be regarded as correlated the correlation time falls to a fraction of a second. The correlation time is more than a curiosity about the μ-DS and can be used as the basis for selecting classifier training data; this selection is discussed in the next section.

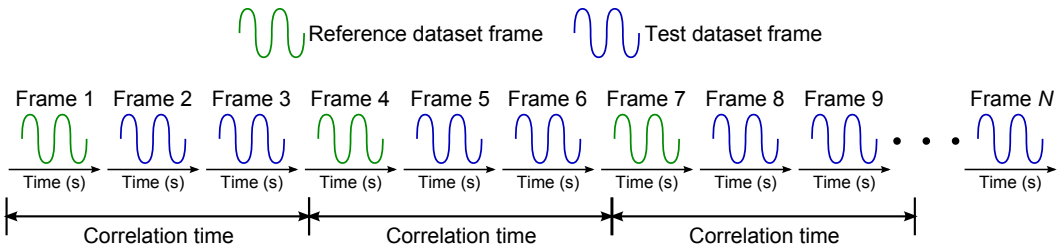


Figure 5.6: The selection of frames for the reference and test dataset.

5.3 Creation Of Datasets

In section 3.1 the two sets of input to the "black box" classifier were defined—the reference and test datasets. The entries for each of the datasets were selected from the framed target data, created above in section 5.2, based on the correlation time. By spacing the frames used to form the reference dataset by the reconstructed signal correlation time the method for creating the datasets became controlled and repeatable. The creation of many data frames from long signals is common in the literature, however the selection of frames for the reference and test dataset is usually random (Stove and Sykes, 2003; Jahangir et al., 2003b; Bilik et al., 2006). While this approach achieves the creation of the required datasets it does not provide insight into how the nature of the reference dataset effects the ATR performance. In the method described here, however, it is possible to consider how varying the correlation between reference and test data effects performance.

Fig. 5.6 shows how correlation time can be used to divide frames into the test and reference datasets. With the frames arranged contiguously, the first frame is always selected for inclusion in the reference dataset. The next frame to be selected is separated from the first by the correlation time. This process is repeated leaving a correlation time between each frame of the reference dataset. The remaining frames are then used for the test dataset.

This process created twenty pairs of reference and test datasets: there were five frame durations and four correlation levels leading to twenty combinations. As would be expected, the two synthesized "unknown" targets were omitted from the reference dataset. The selection of frames from the "unknown" targets for the test datasets wasn't based on the correlation time since the random target decorrelated immediately. Instead a random selection was made with the number of frames selected matching the number of frames available for the tracked vehicle target.

The four correlation levels considered were 0.3, 0.5, 0.7 and 0.9, where these levels represent the value the normalized correlation could fall to before the signal was deemed to have decorrelated. When a low correlation level is used the correlation time is long, see Fig. 5.5, meaning less frames are selected for the reference data. Conversely when the correlation level is high the correlation time is short and the number of frames in the reference dataset is large. It was expected that the larger datasets would result in improved classifier performance since there would be more chance of a strong match between an entry

in the reference dataset and the input being classified. The last factor to consider in the comparison of the reference and test datasets is the observation angles included in each.

The Thales MSTAR data contained examples of each target at several aspect angles; it was decided that these could be used to assess the probability of generalization introduced in section 3.3. Restricting the observation angles included in the reference dataset compared to the test forces the classifier to generalize. If the input is for the target with an observation angle of 30° and there are only head-on (0°) examples in the reference data then the classifier must generalize to identify the target. Reference datasets contain examples at only head-on aspects, at 30° and 150° aspects and at all aspects were created.

All of the processing described above is independent of the ATR approach being taken. On occasion the classifier used placed some requirements on the number of entries in the datasets: for example, k-Nearest Neighbour (k-NN) required sufficient entries in the reference dataset for a fair vote, see section 6.2. When such requirements arise they will be discussed *in situ* rather than attempting to consider them here where they are out of context. In addition to the division of the available data into reference and test datasets further processing is required to simulate the classifier feature extraction stage, see Fig. 3.1. The discussion of this feature extraction is the topic of the next section.

5.4 Classifier Feature Extraction

The "black box" classifier model introduced in section 3.1 described the generalized classification process; the first stage of which was feature extraction. Feature extraction is the process of extracting a feature vector—a data vector suitable for input to the classifier—from the raw data output by the sensor. For an operational ATR system the feature extraction process would be applied to all entries of the reference dataset prior to its use and then to each input as it was presented for recognition. When the classifier is being evaluated all of the inputs are stored in the test dataset and it becomes possible to apply the feature extraction process in advance in the same manner as the reference dataset.

The testing performed with the Thales MSTAR data falls into two broad categories: time domain processing and frequency domain processing. Each of these categories takes a substantially different approach to feature extraction.

5.4.1 Time Domain Processing

In time domain processing the classification is performed on time series data where the samples follow one and other contiguously in time. Each feature vector is a digital representation of an analogue signal. The reconstructed signals formed from the Thales MSTAR data were already in the form so very little feature extraction was required.

The feature extraction was limited to finding the intensity value from the original complex samples.

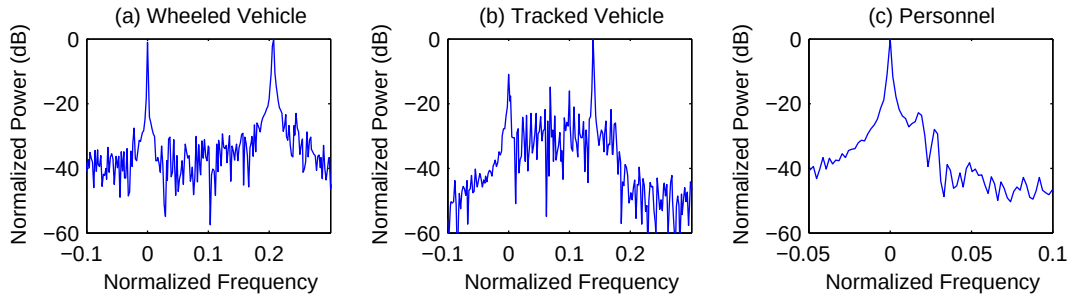


Figure 5.7: The power spectra for the three target classes showing the dominance of the clutter return.

The time domain testing was included to allow investigation of the Dynamic Time Warping (DTW) technique, which was originally developed to classify time series data (Myers et al., 1980). Since each of the data frames created in section 5.2 is a time series the only feature extraction undertaken was to convert the complex samples into real intensity values.

It is possible for DTW to process time series where each sample is a multidimensional feature vector, and the complex samples could have been represented as two dimensional vectors. If the data had been prepared in this manner then it would not have been possible to use it with the k-NN classifier that was used as a benchmark. This will be explained in more detail in Chapter 6 when the classifiers are described. It is also worth noting that a preliminary study demonstrated that there was no significant ATR performance difference when using intensity values compared with complex samples for DTW and k-NN (Smith et al., 2006a).

5.4.2 Frequency Domain Processing

In frequency domain processing the classification is performed on a feature vector derived from the power spectrum of the data. The domain required more sophisticated processing than the time domain where the basic framed data was used as a feature vector. The processing chain had three stages:

1. the removal of ground clutter;
2. frequency normalization; and
3. the reduction of feature vector dimension.

These stages are described below.

5.4.2.1 Removing Ground Clutter

When the Discrete Fourier Transform (DFT) is applied to any of the Thales MSTAR data it is apparent that the most dominant feature is the ground clutter. Fig. 5.7 shows the power spectra for the three target classes as calculated from a 64 ms frame of data. In each spectrum there is a high power response at zero frequency from the ground clutter. For the wheeled vehicle the response was observed as being easily confused with the target response, at frequency 0.2 in the figure; for the personnel spectrum the ground clutter is clearly the most dominant feature; and while the tracked vehicle μ-DS is distinguishable from

the clutter response, clutter is still a major feature of the spectrum. The dominance of the ground clutter in these spectra was increased through the signal reconstruction process when the range bins are summed since clutter power depends in part on the size of the range bin (Stimpson, 1998). When the same data has been used in other studies, where signal reconstruction was not performed, the target signal had comparable or larger amplitude than the clutter (Stove and Sykes, 2003), however it was still a major feature. To ensure that the classifiers were only recognizing the μ-DS and not elements of the ground clutter it was decided to remove the clutter response.

The ground clutter was removed from the signals through a novel variant of the CLEAN algorithm (Högbom, 1974; Hai, 2004) rather than by filtering. Since the MSTAR radar is a ground based system the clutter could be removed with a notch filter centred on 0 frequency. This would have a similar effect to the pulse delay cancellation of section 2.3. But such a treatment ignores the frequency sidelobes of the clutter that could affect the μ-DS. (If the clutter signal consisted solely of a Direct Current (DC) signal then it wouldn't have sidelobes, but since the clutter contains objects such as trees that move in the wind its signal covers a range of frequencies and has sidelobes that could distort the targets' μ-DS.) To remove both the main clutter response and its sidelobes a modified version of the CLEAN algorithm was employed.

The CLEAN algorithm is a technique originally developed in radio astronomy to remove the sampling function from the response of radio telescopes (Högbom, 1974) that has since been adapted for use in many areas, including radar. In radar the algorithm is often used to remove unwanted sidelobes in range profiling. An excellent description of such use is given in (Hai, 2004) in which the algorithm is employed to remove sidelobes from Barker coded waveforms. For use when removing clutter both the main and sidelobes must be removed.

The version of CLEAN used to remove the clutter is provided in algorithm 1. The number of frequency bins, X in the algorithm, was set to correspond to frequencies between $0\,\text{Hz}$ and $\pm 60\,\text{Hz}$. This range was selected by observing several power spectra from the MSTAR data and measuring the extent of the clutter. The clutter was removed from each range bin in turn and the reconstructed signal was formed in the same manner described in section 4.4. Fig. 5.8 shows the power spectra for the three target classes after the clutter has been removed. For the wheeled class all that remains is the target's μ-DS with no significant trace for the ground clutter. There is still a small response at zero frequency for the tracked vehicle, indicating that the clutter was not completely removed. However, there is still significant suppression when the CLEANed power spectrum is contrasted with that of Fig. 5.7. The most dramatic response is for the personnel target class. In Fig. 5.7 it was nearly impossible to see the target μ-DS as it was obscured by the ground clutter sidelobes. In contrast to this Fig. 5.8 shows a clear signature between 0.01 and 0.06 on the normalized frequency scale.

Once the clutter had been CLEANed from the data and the reconstructed signal formed it can be

Algorithm 1 The CLEAN algorithm for ground clutter removal.

```
Repeat for each range bin
  Take the DFT of all pulses in current range bin to find
      frequency spectrum of that bin
  Loop
    Find location of peak amplitude within the first and
        last X frequency bins of the spectrum
    If the peak is greater than the threshold value
      Estimate parameters of the pure tone that would cause
          the peak
      Calculate the frequency spectrum of the pure tone
      Subtract pure tone's spectrum from the original spectrum
          to form the spectrum for the next iteration
    Else
      Break loop
    End if
  End loop
  Perform inverse DFT on the remaining frequency spectrum
      to give clutter free version of the range bin data
End
```

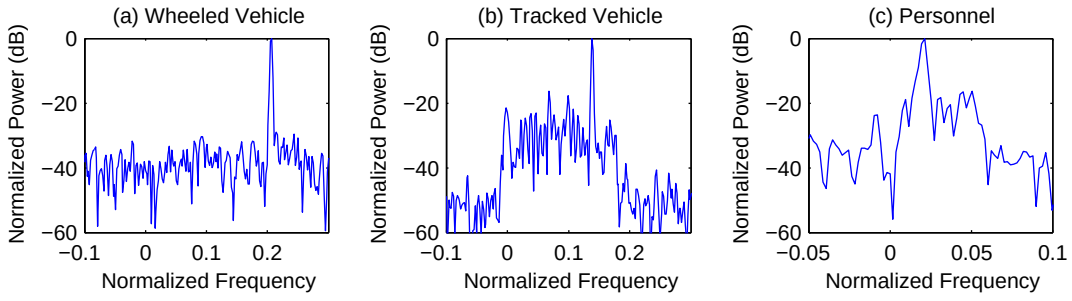


Figure 5.8: The power spectra for the three target classes showing the successful removal of the clutter return.

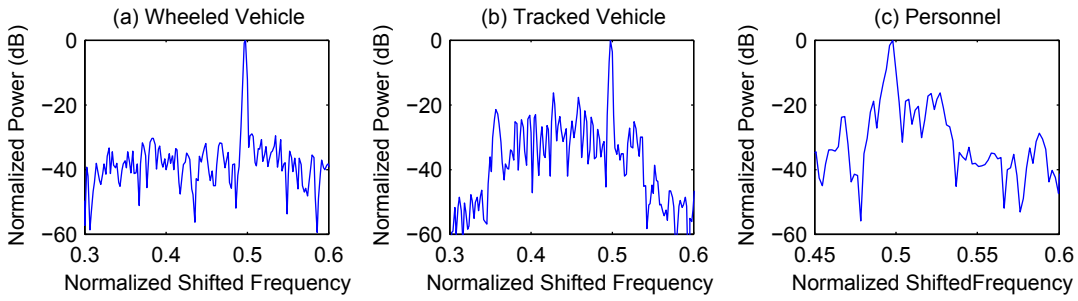


Figure 5.9: The power spectra for the three target classes after frequency normalization by rotational shift.

divided into frames and datasets using the method described in sections 5.2 and 5.3. A further step is applied to work in the frequency domain: the power spectrum for each frame is calculated using the DFT. This is the same as the method that was used to produce the power spectra from the 64 ms frames in Figs 5.7 and 5.8. With the data now converted to a frequency format it can be normalized to remove target velocity information.

5.4.2.2 Performing Frequency Normalization

When observing the power spectra calculated from each data frame it was noted that the variations in target velocity were much more apparent than in the time domain and were a more significant feature than the μ-DS. In Fig. 5.8 the Main Doppler Lines (MDLs) of the targets are seen at normalized frequencies of ≈ 0.20 for the wheeled vehicle, ≈ 0.15 for the tracked vehicle and ≈ 0.02 for the personnel. During the trials when the data was gathered the target velocities were kept as constant as possible. If such data were presented to a classifier to use for both training and testing there would be a high chance the classifier would learn to recognize target speed alone. To prevent this it was decided to normalize the spectra to remove velocity information.

Frequency normalization, to remove velocity information, was achieved by performing a circular shift on the spectra to move the peak value, the MDL, to the central bin. The circular shift operation rotated the elements of given spectrum/feature vector as if the two ends were connected. For example, rotating the sequence $\{1, 2, 3, 4, 5, 6, 7\}$ by 4 results in $\{4, 5, 6, 7, 1, 2, 3\}$. Fig. 5.9 shows the rotation of the frames from Fig. 5.8 to move the MDL to a normalized shifted frequency of 0.5. The normalization goes some way to removing the target velocity information from the spectra, but it is not perfect. The μ-DS bandwidth depends on target velocity and the normalization does not re-scale the bandwidths. However, this was not regarded as problematic since bandwidth is also target descriptor whereas velocity is only a function of velocity of the target.

The normalized power spectra could be considered a reasonable feature vector for use in the classifier. The μ-DS of each target was clearly visible; the dominant ground clutter response had been removed; and the target velocity information had been normalized. Despite these steps there were still

limitations. The dimensionality of the feature vectors was high—for the 64 ms frame there were ≈ 500 elements. There was no guarantee that all element of feature vector were orthogonal leading to dependencies that can limit classifier performance (Duda et al., 2005). Lastly, there are large regions of noise either side of the signatures that provide no target information but serve to increase the feature vector dimension and therefore require more data be available for training. Several of these limitations can be improved by, if not completely overcome by, the application of Principal Component Analysis (PCA) to the datasets.

5.4.2.3 Reduction Of Feature Vector Dimension

PCA provides a mechanism to reduce the impact of the limitations, outlined above, of using the normalized power spectra as feature vectors (Hand and Yu, 2001; Duda et al., 2005; Bishop, 2005). It achieves these improvements by finding a new set of co-ordinates, based on the data covariance matrix, in which to represent the data. In addition to finding the basis vectors of the new space it also provides a score for each basis vector representing its significance. Once the data has been projected into the new space, the elements of each feature vector that correspond to the low score basis vectors can be discarded. This rejection of some elements reduces feature vector dimensionality as well as removing the noise elements of the original vectors since it is the noise elements that result in the low scored basis vectors. The orthogonality of the elements of the new feature vectors is also improved, although there is still no guarantee that they are completely orthogonal since the PCA process is based on the group covariance matrix rather than the covariance matrix of each class (see the e-mail from the author of (Hand and Yu, 2001) in Appendix A). The operation of PCA may be explained pictorially.

Fig. 5.10 shows the PCA process for simulated data from two classes where each data sample has two elements, x_1 and x_2. In part (a) of the figure the data has been plotted in the measured co-ordinate space and the principal components have been superimposed on top of this plot as the red and blue lines. In this data it is clear that there is a dependency between x_1 and x_2 since the greater the value of x_1 the greater the value of x_2. In part (b) of the figure the data has been projected onto the principal components, P_1 and P_2, thus making them the basis vectors of the new feature space. The two basis vectors, P_1 and P_2, no longer depend on each other since the position of a data point along the P_1 axis does not influence its position on the P_2 axis. In part (c) the data has been projected onto P_1 alone reducing the dimensionality of the data. It is clear that the two classes are still separable in this single dimension space making it more suitable as the input to a classifier since the lower dimensionality will require less reference data to find the decision boundary. The PCA processes is based around eigen decomposition.

The principal components, or new basis vectors, shown in Fig. 5.10 are generated by solving the

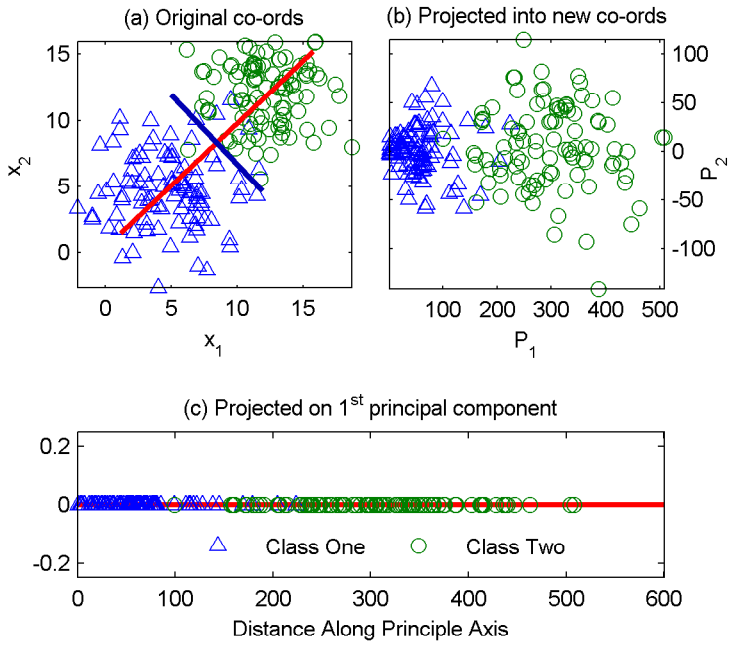


Figure 5.10: The process for performing Principal Component Analysis (PCA).

eigen equation (Duda et al., 2005):

$$\mathbf{C}\mathbf{p} = \lambda\mathbf{p} \tag{5.9}$$

where: $\mathbf{C}$ is the data covariance matrix defined as:

$$\mathbf{C} = \left\langle (\mathbf{x}_k - \langle \mathbf{x} \rangle)(\mathbf{x}_k - \langle \mathbf{x} \rangle)^{\mathrm{T}} \right\rangle_k \tag{5.10}$$

and $\mathbf{x}_k$ is the k^{th} data sample; $\mathbf{p}$ is an eigenvector; and λ an eigenvalue. There will typically be several solutions to (5.9) and each eigenvector represents a different principal component; the solution with the largest eigenvalue is the most important principal component, the second largest eigenvector, the second most important and so on. The sum of the eigenvalues gives an indication of the total variance in the data making it possible to calculate the percentage of total variance along each principal component. This knowledge is used to reduce the dimensionality of the data: if there are, say, fifty principal components yet it is apparent that 99% of variance is accounted for in the first ten then the remaining forty can be rejected. The formulation of PCA given here provides the maximum variance in the projected data (Bishop, 2005) and is useful in the development of classifiers[1]. If the data is projected onto the first three principal components it is possible to plot it to review the class separation.

Initial attempts to apply PCA to the Thales MSTAR data gave poor results. The eigenvalues in the solutions of (5.9) suggested that nearly 100% of the data variance was along the first eigenvector and

[1]It should be noted that there is an alternative formulation of PCA based on minimizing the distortion between the original feature space and the reduced dimension space that is described in (Bishop, 2005). The result is the same, but it is refreshing to know that the projected data has maximum variance and minimum distortion.

Scale	Five Random Amplitude Samples					Total Standard Deviation
Linear	1.4×10^4	3.2×10^4	2.6×10^4	3.7×10^4	5.6×10^4	4.3×10^4
Log	41.5	45.0	44.2	45.7	47.5	2.2

Table 5.2: Variation of MDL amplitude for a tracked vehicle.

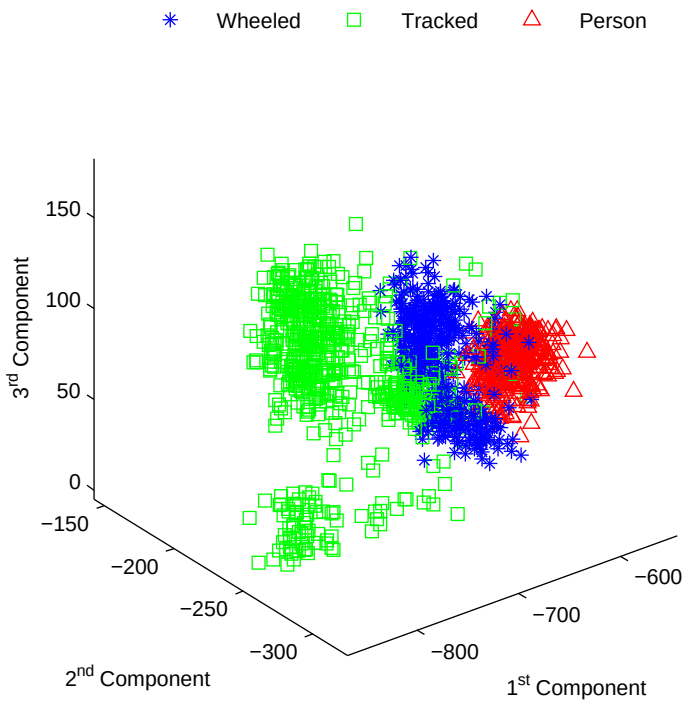


Figure 5.11: The first three principal components of real μ-DS data showing the separation of the classes.

that the others were redundant. This result was regarded as suspicious since the original spectra clearly displayed significant features across a number of frequency bins. Closer analysis also showed that the eigenvector in the direction of maximum variance tended to coincide with the position of the MDL in the normalized spectra. The amplitude scale on the normalized power spectra had been linear and it was realized that due to both the different RCS of the targets and RCS scintillation effects (Skolnik, 1990) this caused there to be enormous variation in the amplitude of the MDL. Table 5.2 contrasts the amplitudes of the MDL for some 64 ms feature vectors for a tracked vehicle when linear and decibel scales are used. For the linear scale the standard deviation, the square root of the variance, is very large, or order 10^4, but for the logarithmic scale it is just 2.2. Using a decibel scale for the power spectra gave a greatly improved PCA performance and it became possible to use the technique to successfully process the MSTAR data.

Fig. 5.11 shows a plot of the first three principal components for the target classes calculated from a 64 ms data frame. There are three distinct "clouds" of data points in the feature space: one for tracked vehicles, one for wheeled and the third for personnel. That the classes fall into these distinct regions is promising from a classification perspective as it suggests they can be discriminated between. Unfortu-

nately the figure also shows that there is some degree of overlap between the data clouds. Predicting the proportion of inputs that will be misclassified as a result of this overlap was considered in section 3.6. Production of a comparison figure for the data before the application of PCA is not possible. Prior to the dimension reduction the feature vectors had over 500 elements each and it is not possible to plot such high dimension vectors on a two-dimensional page.

This chapter has defined the pre-processing stages of the classifier and explained how additional targets were synthesized. With these pre-processes complete it is possible to begin to discuss the classification of the data and its implications for radar ATR. The next two chapters introduce the classifiers used in the studies, explain the tests that were conducted and consider the implication of the results. The chapters are divided into time domain processing and frequency domain processing.

5.5 Summary

This chapter described the pre-processing of the Thales MSTAR data used in the study. The pre-processing was broken into two stages. The first stage converted the data, which was gathered on a field trial, to be more like the data an operation radar would output. The second stage represented the feature extraction the radar ATR system would perform ahead of the classifier. Before describing the pre-processing, however, the method by which additional target classes were synthesized was described.

The Thales MSTAR dataset contained just three classes. In order to test an ATR solution's "unknown" detection capability there must be some classes in the test dataset that are not in the reference dataset. Two additional classes were simulated to use as "unknowns". The first was a high amplitude random signal selected from a uniform distribution and the second a simulation of a propeller. To increase their realism, the synthesized signals were added to clutter measurements from the trial where the data was gathered. The resulting data was saved in the same format as the radar data so that it could be subject to the same pre-processing.

To simulate the dwell times of a scanning radar, the long data files from the Thales MSTAR dataset were divided into short frames. Because the Thales data had been gathered on a trial, the durations of the signals were much longer than those that would result from an operation radar. To increase realism, the long files were cut into short frames. Some of the frames were selected for inclusion in the reference dataset of the classifier with the remainder used for the test dataset. The selection process was based on the correlation time of the data. Using this metric provided a quantifiable method for reference dataset selection.

The feature extraction process was to convert the time domain data frames into power spectra. Once converted, the data was normalized to remove target velocity information. The normalized frames then had PCA applied to them to reduce their dimensionality. Analysis of the first three principal components showed the data to be divided into clearly defined regions of feature space. Such division suggested that

automatic recognition would be possible.

Chapter 6

Monoperspective Time Domain Classification

The analysis of Micro-Doppler Signature (μ-DS) radar Automatic Target Recognition (ATR) has been broken into two parts the first of which, time domain classification, is discussed in this chapter. As part of the discussion the Dynamic Time Warping (DTW) method of classification is introduced and its significance to radar ATR explained. The k-Nearest Neighbour (k-NN) classifier that is used as a benchmark is also described. It is observed that the use of k-NN places further requirements on the formation of the reference datasets compared to the method of section 5.3. A modified version of the dataset creation algorithm is developed to cope with these requirements.

The testing strategy employed results in a series of smaller investigations. First, consideration is give to the effect of using intensity data, rather than coherent samples. The impact of the number of target aspects included in the reference dataset is then explored. Finally, the DTW classifier alone is used to investigate the pure correlation time method—where no allowance is made for the requirements of k-NN—of creating reference datasets.

To close the chapter the results from all the investigations are compared.

The ability of battle field radar operators to recognize the Micro-Doppler Signature (μ-DS) they hear output by their equipment (Stove and Sykes, 2003; Bilik et al., 2006) suggested that techniques used in speech recognition may be of value in radar Automatic Target Recognition (ATR). The speech signal, like the output of the battlefield radar, is an audio signal that encodes information and the process of extracting meaning from it involves identifying the words being spoken. Dynamic Time Warping (DTW) is an early speech recognition technique that was identified as being particularly suitable for recognizing the μ-DS since it is robust to some of difficulties anticipated for such classification. The technique can operate on data in the time domain, greatly simplifying data pre-processing since the samples recorded by the radar can be used directly. DTW also makes an ideal starting point for an ATR investigation since it is template matching technique and so relatively simple to implement.

Template matching methods rely on the direct contrasting of the data under test with a reference

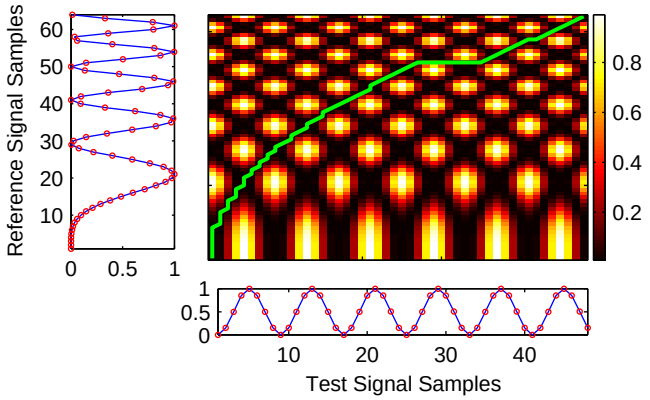


Figure 6.1: The operation of the DTW algorithm.

dataset. The most commonly used template matching method in k-Nearest Neighbour (k-NN) so it was decided to use this method as a benchmark to compare DTW with. Unlike DTW, k-NN is not robust to the difficulties anticipated in μ-DS ATR; it was expected that DTW would perform much better. The first two sections of this chapter will describe the two template matching algorithms. The remainder of the chapter will be dedicated to explaining the testing strategy and presenting the results.

6.1 Dynamic Time Warping

DTW is a template based classification technique from speech recognition that is able to compensate for displacements between features in a test and reference signal; such displacements occur when signal are generated at different rates. The variation in frequency of the radar echo signal due to the Doppler shift may also be considered a mechanism that displaces features in a signal: if the same target is observed travelling at different velocities then both the time and frequency domain version of its signal will differ between the observations. The comparison metric is the Global Distance (GD) between the signals, and this metric is commonly normalized to allow signals of different lengths to be compared. The normalizing constant depends on the specific implementation of the DTW algorithm although it is normally a linear combination of the lengths of the signals under comparison. Classification is achieved by comparing an input with a pre-classified reference dataset. The declared class is the same as the reference entry resulting in the lowest Normalized Global Distance (NGD). A detailed description of the DTW algorithm is given in (Sakoe and Chiba, 1978; Myers et al., 1980; Sankoff and Kruskal, 1999). Here an abridged description is provided to allow sufficient appreciation to understand DTW's use in this investigation.

The operation of DTW is shown in Fig. 6.1. The test and reference signals are shown to the bottom and left of the figure respectively and are compared in a multi-stage process. First, a local distance map is calculated that details the distances between all the samples in both signals, this is shown as the intensity plot in the figure. Secondly, the warping path is found by traversing the local distance map from the start

to the end of both signals (the bottom left corner to the top right of the map). It is represented by the line superimposed on the map in the figure. Thirdly, the GD between the two signals is found by summing the local distances along the warping path. Lastly, normalization of the GD is performed. Even after normalization the NGD obtained will depend on the warping path used leading to the question, "Which is the optimum warping path?"

In order to define the optimization process it is necessary to formalize the description of DTW. The test and reference signals, $T[j]$ and $R[k]$, are series of samples where each sample may be a vector of measurements:

$$T[j] = \{\mathbf{t}[1], \mathbf{t}[2], \mathbf{t}[3] \dots \mathbf{t}[N_T]\} \; = \; \{\mathbf{t}[j]\} \quad \text{where} \quad \{j \in \mathbb{N} : j \leq N_T\} \tag{6.1}$$

and

$$R[k] = \{\mathbf{r}[1], \mathbf{r}[2], \mathbf{r}[3] \dots \mathbf{r}[N_R]\} \; = \; \{\mathbf{r}[k]\} \quad \text{where} \quad \{k \in \mathbb{N} : k \leq N_T\} \tag{6.2}$$

The warping path, $F[n]$, is then a series of co-ordinates that describe the route taken across the local distance map and represents the warping function that approximately realizes a mapping from the time axis of the test series onto the reference series (Sakoe and Chiba, 1978):

$$F[n] = \{f(1), f(2), f(3) \dots f(N_\mathrm{f})\} = \{f(n)\} \quad \text{where} \quad \{n \in \mathbb{N} : n \leq N_\mathrm{F}\} \tag{6.3}$$

in which

$$f(n) = (j(n), k(n)) \tag{6.4}$$

is a step in the path and

$$N_\mathrm{f} \leq N_\mathrm{R} + N_\mathrm{T} \tag{6.5}$$

is the maximum path length. General constraints are placed on the first and last steps of $F[n]$ to ensure the warping path starts at the beginning of the signals and finishes at their ends: $f(1) = (1, 1)$ and $f(N_\mathrm{F}) = (N_\mathrm{T}, N_\mathrm{R})$. Further constraints, known as local constraints, are placed on the progression from $f(n-1)$ to $f(n)$ to limit the available warping paths. The exact nature of these constraints is not required until later in the description of DTW and so are ignored for the moment. The last item to be formalized prior to defining the optimization is the GD to a point on the warping path, $GD(F[n])$.

$$GD(F[n]) = \sum_{p=1}^{n} d\left(f(p)\right) \omega[p] \tag{6.6}$$

where $\omega[p]$ is the weighting associated with the p^{th} arc of the warping path and $d\left(f(p)\right)$ is the local distance between $\mathbf{t}[j(p)]$ and $\mathbf{r}[k(p)]$, typically $d\left(f(p)\right)$ is the Euclidean distance i.e. $\|\mathbf{t}[j(p)] - \mathbf{r}[k(p)]\|$.

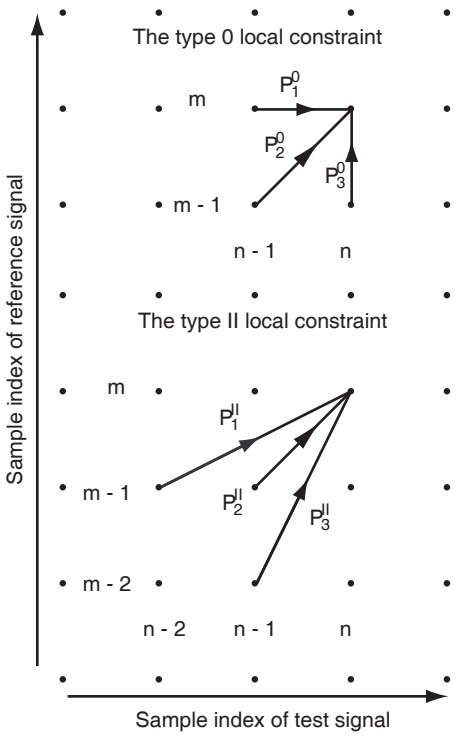


Figure 6.2: The two possible local constraints used by the DTW algorithm in this investigation.

The local distance map that has been discussed previously is simply a two dimensional array holding the value of $d\left(\left(j(n), k(n)\right)\right)$ for all n. It is clear that the value of $GD\left(F[n]\right)$ depends on the number of steps in the warping path, this is N_{F} for the last point of the path. It is therefore necessary to normalize (6.6) giving the NGD, $D\left(F[n]\right)$:

$$D\left(F[n]\right) = \frac{\sum_{p=1}^{n} d\left(f(p)\right)\omega[p]}{\sum_{p=1}^{n}\omega[p]}. \tag{6.7}$$

Using (6.7) it is possible to define the optimum warping path as: *the optimum warping path is that which results in the smallest distance between the two signal and therefore minimizes* $D\left(F[N_{\mathrm{F}}]\right)$. If the minimum distance is denoted by $\hat{D}$ then $\hat{D} = \min\limits_{F[n]}\left\{D\left(F[N_{\mathrm{F}}]\right)\right\}$. To achieve this minimization over the warping path the final part of the DTW algorithm must be introduced: the local constraint.

As mentioned, the local constraints limit the progression from $f(n-1)$ to $f(N)$: they describe the steps that may be taken in forming a warping path. There are many types of local constraints, identified with numeric designations in DTW literature. This investigation used type 0 and type II local constraints. Type 0 constraints were chosen as they allow the maximum degree of warping while type II constraints were selected as they require the least number of computations, of all the standard constraint types, resulting in fastest execution of the algorithm (Myers et al., 1980). The two types of constraint are shown graphically in Fig. 6.2. They can also be described numerically:

$$f(n-1) = \begin{cases} (j(n-1), k(n)) \\ (j(n-1), k(n-1)) \\ (j(n), k(n-1)) \end{cases} \tag{6.8}$$

for the type 0 constraints; and

$$f(n-1) = \begin{cases} (j(n-2), k(n-1)) \\ (j(n-1), k(n-1)) \\ (j(n-1), k(n-2)) \end{cases} \tag{6.9}$$

for the type II constraints. Associated with each local constraint is a weighting that provides the $\omega[n]$ from (6.6).

Because the length of arc covered by each local constraint may vary a weighting is applied to prevent the variation in length affecting the probability of selection for use in the warping path. Since reducing the number of steps in the warping path will minimize $\hat{D}$ using longer arcs would be desirable even if they resulted in a sub optimal warping. For the two local constraints used the weightings are:

$$\omega[n] = 1, \tag{6.10}$$

i.e. there is no weighting for any constraint, for the Type 0 constraints, and

$$\omega[n] = \begin{cases} 3 \quad \text{if} \begin{cases} j(n) - j(n-1) = 2 \quad \text{and} \\ k(n) - k(n-1) = 0 \end{cases} \\ 2 \quad \text{if} \begin{cases} j(n) - j(n-1) = 1 \quad \text{and} \\ k(n) - k(n-1) = 1 \end{cases} \\ 3 \quad \text{if} \begin{cases} j(n) - j(n-1) = 1 \quad \text{and} \\ k(n) - k(n-1) = 2 \end{cases} \end{cases} \tag{6.11}$$

for the Type II constraints.

The comparison metric between the two series is the NGD, $D\left(F(N_{\mathrm{F}})\right)$, that has the term $\sum_{p=1}^{n} \omega[p]$ in its denominator. The calculation of the denominator can prove problematic when recursive or dynamic programming techniques are employed. Fortunately it has been shown that the expression can be simplified to a constant term without compromising the calculation of the NGD (Myers et al., 1980):

$$\sum_{p=1}^{n} \omega[p] \approx \begin{cases} N_{\mathrm{R}} & \text{for type 0} \\ N_{\mathrm{R}} + N_{\mathrm{T}} & \text{for type II} \end{cases} \tag{6.12}$$

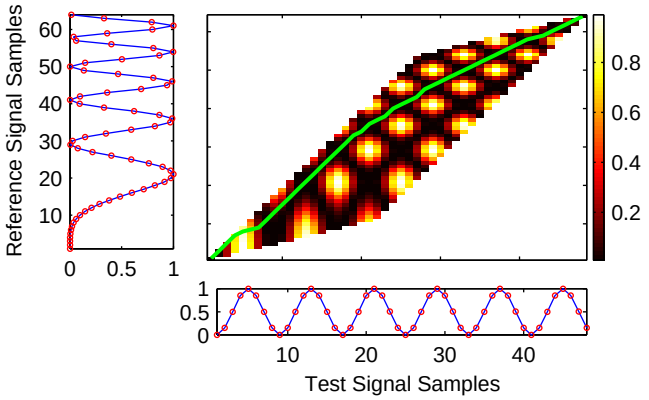


Figure 6.3: The operation of the DTW algorithm using type II local constraints.

It is now possible to develop a dynamic programming routine to evaluate the optimum NGD, $\hat{D}$, between the test and reference series based on (6.3) to (6.12). The initialization step of the algorithm is the GD at $F[1]$:

$$g_1\left(f(1)\right) = d\left(f(1)\right)\omega[1]; \tag{6.13}$$

and the iterative step of the algorithm is the minimized NGD to step $F[n]$ of the warping path:

$$g_n\left(f(n)\right) = \min_{f(n-1)}\left\{g_{n-1}\left(f(n-1)\right) + d\left(f(n)\right)\omega[n]\right\}. \tag{6.14}$$

The NGD, $\hat{D}$, is then equal to:

$$\hat{D} = \min_{F[n]}\left\{D\left(F[N_{\mathrm{F}}]\right)\right\} = \begin{cases} \dfrac{1}{N_{\mathrm{R}}} g_n\left(f(N_{\mathrm{F}})\right) & \text{for type 0} \\ \dfrac{1}{N_{\mathrm{R}} + N_{\mathrm{T}}} g_n\left(f(N_{\mathrm{F}})\right) & \text{for type II} \end{cases} \tag{6.15}$$

remembering that the weighting function $\omega[n]$ will be different for the two types of local constraint. If a series is compared with itself then the optimum NGD is 0. The more different the two signals the greater the NGD becomes.

The example of DTW operation shown in Fig. 6.1 uses the type 0 local constraints. The same signals are compared in Fig. 6.3, but this time using type II local constraints. The most striking change between the two sets of local constraint is the local distance map shape. It is a parallelogram for type II constraints, known as the Itakura parallelogram, containing fewer points than the rectangle of the type 0 example. It is the reduction in number of points that results in type II constraints requiring fewer calculations. The warping path chosen is also different since it is constrained to be within the limits of the Itakura parallelogram.

A series of echo pulses are received from a target during a radar dwell may be considered the samples of a signal. Such signals will contain the μ-DS as described in section 4.1. Using DTW the

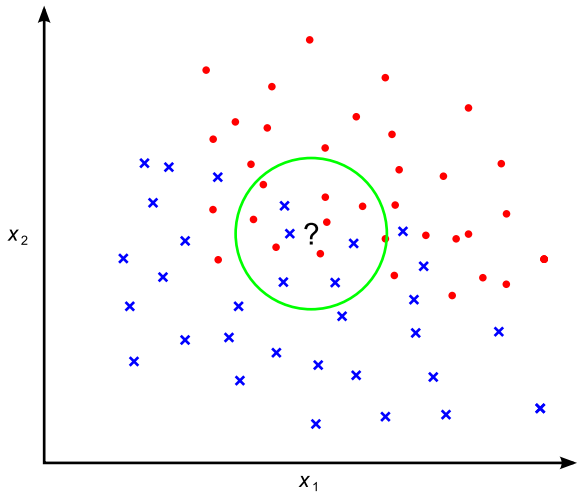


Figure 6.4: The operation of the k-NN algorithm.

signals, and hence μ-DS, of a radar target may be compared with a pre-classified reference dataset to determine the class of target being illuminated. The nature of the DTW algorithm does not limit the individual samples of a signal to be scalar so it is possible to compare signals with complex samples. Although it might be expected that the extra phase information would assist target recognition complex samples give no significant performance increase, compared to intensity values, when DTW was used for radar ATR (Smith et al., 2006b). The increase in computation required to process two numbers for each sample actually makes complex sample DTW much slower than the intensity version; as such only intensity based DTW was used for the majority of the investigations for this work. The results from (Smith et al., 2006b) relating to the use of intensity rather than complex DTW are repeated at the start of section 6.4.

6.2 k-Nearest Neighbours

The k-NN classifier is one of the most common non-parametric classification techniques and is well described in the literature (Duda et al., 2005). The working of the algorithm is similar to DTW since each input is compared with the content of the reference dataset to see which class it is most likely to come from. Unlike DTW the comparison is linear and there is no warping capability to better align similar features. The operation of k-NN is shown in Fig. 6.4.

k-NN classifies an input by finding the most common class in the k closest entries to the input in the reference dataset. In Fig. 6.4 there are two classes, red circles and blue crosses. The input is at the position indicated by the black question mark and the thirteen closest neighbours are those inside the green circle. Although it is possible to use other distance metrics the Euclidean distance is normally used to find the nearest neighbours. Classification is then achieved by performing a simple vote among the neighbours to the input: the declared class is whichever is most prevalent among the neighbours. In

the figure the input would be declared as a red circle since there are eight of them among the neighbours and only five blue crosses.

The measure of "closeness" used in the k-NN classifier is commonly the Euclidean distance. This involves treating both the reference and test series as vectors and places a limit on the nature of the data that can be compared. The measurement represented by each sample of the series must be a scalar so that the overall series may be considered as a vector. This is a substantial limitation compared with DTW where each sample may be a vector representing several measurements. When processing time domain radar data this requirement means that only real valued signals can be used removing the ability to distinguish the sign of the Doppler shift. A second difference between k-NN and DTW classifiers is the necessity of an offline training stage when using k-NN.

Clearly the value of k will greatly affect the performance of the k-NN classifier and its value must be selected before the classifier can be used. The optimum value of k can be estimated from the reference dataset using hold one back cross-validation (Duda et al., 2005). For every candidate value of k each entry of the reference data is selected in turn and classified against the rest of the data. By keeping record of the results it is possible to determine which values of k resulted in the most number of correct classification. Selection of candidate values for k is dependent on the number of classes being discriminated between and the size of the datasets. So that the majority vote remains fair k must be equal to the number of classes plus one. Furthermore, there must be at least k entries for each class in the reference dataset. It is also useful for k to be an odd number to prevent their being a tie in the vote. In this work k was ranged between 5 and 19 with the upper limit being selected to ensure that when longer frame durations were used, and there were less frames to divide between the test and reference dataset, the test dataset still held enough entries for a meaningful test.

Although simple, in the limiting case the k-NN classifier's performance tends to the best possible (Duda et al., 2005). Despite this it was expected that DTW would outperform k-NN when classifying the μ-DS. Because the comparison in k-NN is linear, a small change in signal phase could mean that the Euclidean distance between an input and examples of the same class in the reference dataset would be large. This large distance measure could then allow pre-classified examples for classes that are a poor match to the input to influence the majority vote. To overcome this deficiency, and reach the limiting case where the k-NN classifier becomes optimal, would require a reference dataset of impractically large size since every possible phase variation would require an entry in the reference dataset.

6.3 Testing Strategy

The objective for performing micro-Doppler signature classification in the time domain was to investigate whether DTW was a suitable method for radar ATR. Secondly the testing provided an opportunity to examine whether it was necessary to perform pre-processing on the raw radar data. The use of k-NN

as a benchmark classifier was particularly valuable in assessing the secondary objective since DTW is specifically intended for use with time varying data while k-NN is not. To complete the investigation, four sets of tests were conducted.

The first set of tests investigated the significance of using complex, I- and Q-channel, data compared with intensity data for μ-DS classification. This testing could only be conducted using the DTW algorithm since it is not possible to operate k-NN with complex values. Simplified datasets were used for these tests: only target data from head-on Line Of Sights (LOSs) were used and there were no variant or unknown classes. Further simplifications included the disabling of the "unknown" threshold in the classifier. For the complex datasets each sample of the data frame was a two element vector. The first element held the real part and the second the imaginary part of the original complex sample. For the intensity data each sample of the data frames held the sum of the squares of the two parts. The data sets used in this testing are referred to as the *simplified intensity test datasets*.

The second set of tests were a full investigation of DTW using k-NN as a benchmark. The reference datasets only included examples of the three known targets travelling head-on to the radar, at 0° from the LOS, while the test dataset included examples from all available angles as well as entries for the synthesized "unknown" targets. By considering the test dataset entries from non 0° to LOS angles as variants of the examples in the reference dataset it was possible to test the classifiers' generalization capabilities. The inclusion of the "unknown" targets permitted the examination of the effect of the "unknown" threshold. To set the threshold the classifier was run against each reference dataset with the appropriate test dataset to the point where the distances had been calculated. The minimum, average and standard deviation of the distances was found and thirty exponentially spaced "unknown" thresholds tried in the interval minimum-distance to average-distance-plus-one-standard-deviation. The limits of the threshold intervals are shown in Table 6.3. The selection of the best threshold was governed by the criteria of $P_{\text{cc}} = 0.8$ that is considered to be an acceptable level for real world use (Stove, 2006). If one of the thirty thresholds used did not provide a suitable value of P_{cc} then linear interpolation was used to find a value for the threshold that would give $P_{\text{cc}} \approx 0.8$. Rather than set an individual threshold for each frame-duration correlation-level pair, effectively defining a new classifier for each, the threshold was set based on whichever pair gave the best classifier performance, i.e. the two types of Receiver Operating Characteristic (ROC) curve were produced for each frame-duration correlation-level pair and the pair with best performance was used to set the threshold for that reference dataset. The use of the test dataset in setting the "unknown" threshold is less than ideal, however, there was insufficient data available to take an alternative approach.

In order that the k-NN classifier could be used with the data, further constraints had to be applied to the formation of the datasets. The basic procedure for selecting frames for the reference and test datasets was described in section 5.3 but this did not allow for the extra requirements of k-NN, described in

Table 6.1: Summary of datasets for time domain classification.

Name	Aspects In Reference Dataset	Aspects In Test Dataset
Simplified Intensity Test Datasets	Head-on (0°to LOS)	Head-on
Single-Aspect Datasets	Head-on	All available
Two-Aspects Datasets	30° and 150° to LOS for vehicles 45° and 135° to LOS for personnel	All

section 6.2, that give a minimum number of entries for each class that must be included in the reference dataset. The time between frames in the reference dataset was therefore changed from being the average correlation time to whichever was the smallest of the average correlation time and the time required to give enough entries for k-NN to operate fairly. This process was repeated for each of the intended correlation levels and the times calculated were still referred to by the correlation levels that gave rise to the original average decorrelation distance. Since the reference dataset included examples of the targets from a single aspect, these dataset are referred to as the *single-aspect datasets*.

The third set of tests were very similar to the second, but included two, rather than one, aspects of the known targets in the reference dataset. The intention of the tests was to see how increasing the information content of the reference dataset affected the classifier performance. The aspects used in the reference dataset were 30° and 150° to the radar LOS for the vehicles and 45° and 135° to the radar LOS for the personnel targets. The same procedures as the single-aspect datasets were used for setting the "unknown" threshold and dealing with the requirements of k-NN. These datasets are referred to as the *two-aspect datasets*.

The last set of tests conducted sought to investigate the correlation time method for selecting reference data. A pure version of the method for reference dataset creation described in section 5.3 was used. The resulting datasets were only used with DTW as they did not meet the requirements for k-NN. These datasets are referred to as the *pure correlation time datasets*. The dataset details are summarized in Table 6.1.

The following results and discussion are broken into the following sections:

- the use of intensity, rather than complex, data;
- dataset size;
- setting of the "unknown" threshold; and
- evaluation of classifier performance

6.4 Results And Discussion

6.4.1 Using Intensity Data

To investigate the difference between using coherent and intensity data a single comparison metric was used: the probability of correct classification, P_{cc}. The variation of P_{cc}, averaged over the three classes,

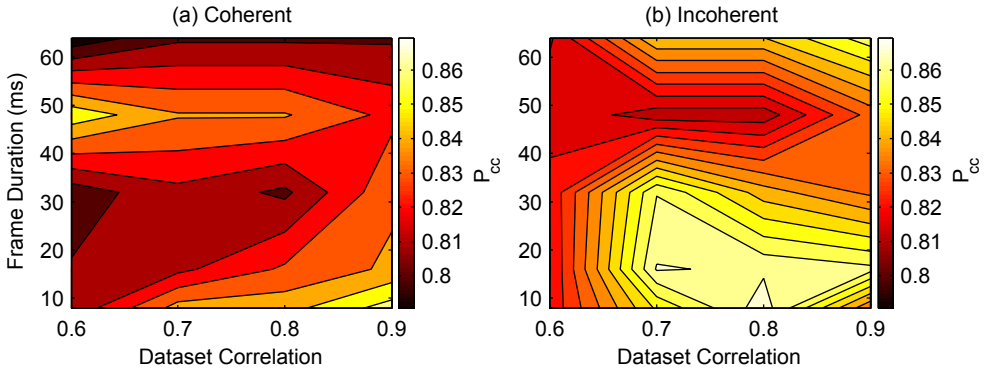


Figure 6.5: Comparison of average P_{cc} for coherent and intensity datasets.

with frame duration and reference dataset correlation level is shown in Fig. 6.5. The intensity scale is the same in the two parts of the figure, [0.79, 0.87]. With such a limited range the fluctuations in P_{cc} tend to overwhelm the general trends in the data. Detailed analysis (Smith et al., 2006b,c), and the results for other datasets discussed below, has shown that longer frames and higher reference dataset correlation level result in better P_{cc}. For these tests, however, these variations are of less interest than the difference between the two figures.

Allowing for the limited intensity range amplifying fluctuations in the results, there is little difference between the coherent samples and the intensity samples datasets. This result is somewhat surprising, since coherent radar data is considered to contain more information (it is possible to determine the sign of a Doppler shift) but it is fortunate in this situation. In the remainder of the testing conducted intensity data, rather than coherent data, is used as the k-NN classifier is able to process it. Knowing that the results obtained by DTW when operating with intensity data are the same as those for coherent data therefore increases confidence in the work presented.

Before continuing with the analysis of classifier performance the impact of the reference dataset creation process on the datasets and the setting of the "unknown" threshold will be considered.

6.4.2 Reference And Test Dataset Creation

Table 6.2 shows the sizes of the reference datasets generated for each frame-duration dataset-correlation pair grouped by dataset name (the simplified intensity test dataset is not included as it was for a specific test and not directly comparable with the others). Short frame durations and high levels of correlation resulted in more entries. Correlation level affects the reference dataset size. It is an indicator of how much variation is considered acceptable between same class entries in the reference dataset. When the differences become unacceptably large, i.e. the entries are decorrelated, a new entry is added to the reference dataset. It follows that the less difference there is between reference dataset entries (higher correlation), the larger the reference dataset will be. Since all the frames for each class were cut from a single reconstructed signal the dataset correlation is effectively setting the size-of-reference-dataset to

size-of-test-dataset ratio. Meanwhile, frame duration affects the total number of frames available. The shorter the frame duration the more frames can be cut from the reconstructed signal so more frames were available for both the reference and test datasets.

The classifier intended for use with the reference dataset also affects the size. Datasets for use with the k-NN algorithm are larger since k-NN required minimum numbers of entry per class for proper operation (Duda et al., 2005). As seen in Table 6.2 the pure decorrelation time dataset that is only used with DTW, had the least entries.

The more aspects held in the reference dataset the larger it will be. Table 6.2 shows the two-aspect datasets as approximately twice the size of the single-aspect ones. Keeping the reference dataset size to a minimum is desirable since during classification the input signal is compared with each entry. Smaller datasets therefore lead to faster classifications. The algorithm for reference dataset creation, described in section 5.3 and modified in section 6.3 considers the data for each target aspect independently. Entries for each aspect are added based solely on the signal correlation time. This is a naïve approach that ignores the potential gain in information about a target that results from having multiple aspects included in the reference dataset. It will be seen in the results that come later in this chapter that some classifiers, such as DTW, are able to generalize from the data in the reference dataset (see the results for P_{Gen} in figures 6.10 and 6.11 and Table 6.4). Allowing for this generalization in the creation of the datasets would permit a reduced number of each target aspect, relative to the number required by the correlation time, to be included in the reference dataset. As the number of included aspects increased the classifier could be assumed to generalize from each aspect and so overcome the drop in information associated with having a reduced number of any one aspect.

6.4.3 Setting Of The "Unknown" Threshold

The ROC curves used to set the "unknown" threshold for classifiers operating with the single-aspect datasets are shown in figures 6.6 to 6.8. In each figure the curves for the different frame durations are shown while the dataset correlation is held constant. Common across all type (a) ROC curves was the failure of P_{Dec} to reach 1 and the step at $P_{\text{FA}} \approx 0.5$, which was most pronounced in the type II DTW curves. Table 6.3 details the threshold settings for all reference datasets and classifiers and the frame-duration dataset-correlation pairing upon which the setting was based.

Setting the unknown thresholds of Table 6.3 required some intuitive steps to be taken. In contrast to the original intention of interpolating between thresholds to meet the criteria $P_{\text{cc}} = 0.8$ it was required to assess each ROC curve to determine how best to set the threshold. Some were set using interpolation, these are indicated by a dagger (†) in the table, with the remainder set after considering the variation of the type (a) ROC curve. For many there was a clear step at $P_{\text{FA}} = 0.5$, Fig. 6.8 shows this most dramatically. In these situations the threshold was set to give the maximum value of P_{Dec} for $P_{\text{FA}} = 0.5$

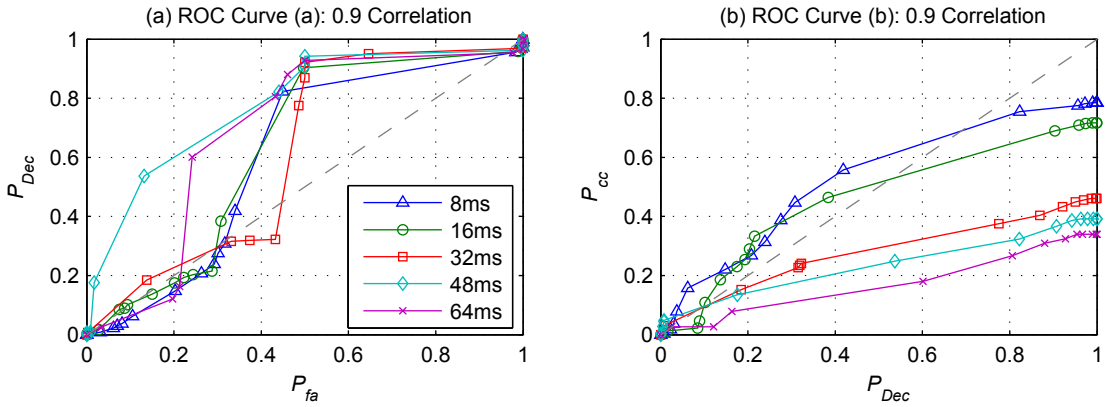


Figure 6.6: The type (a) and type (b) ROC curves for the k-NN classifier.

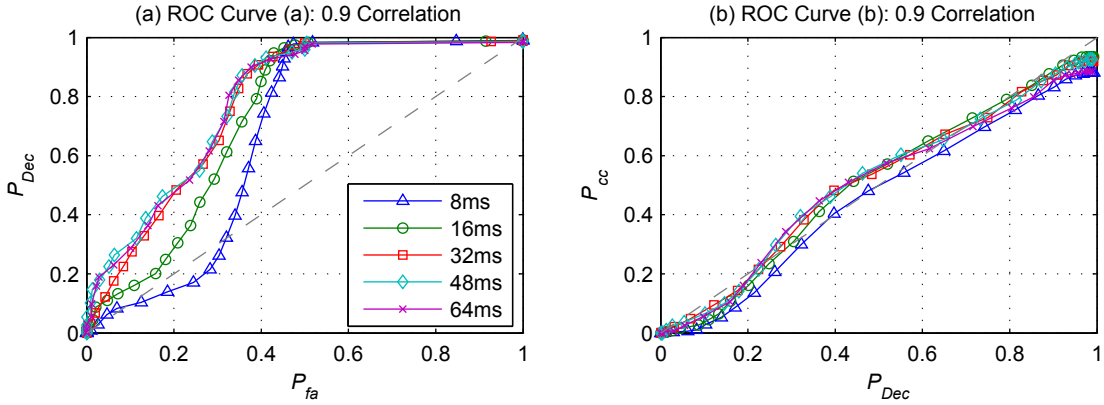


Figure 6.7: The type (a) and type (b) ROC curves for the type 0 DTW classifier.

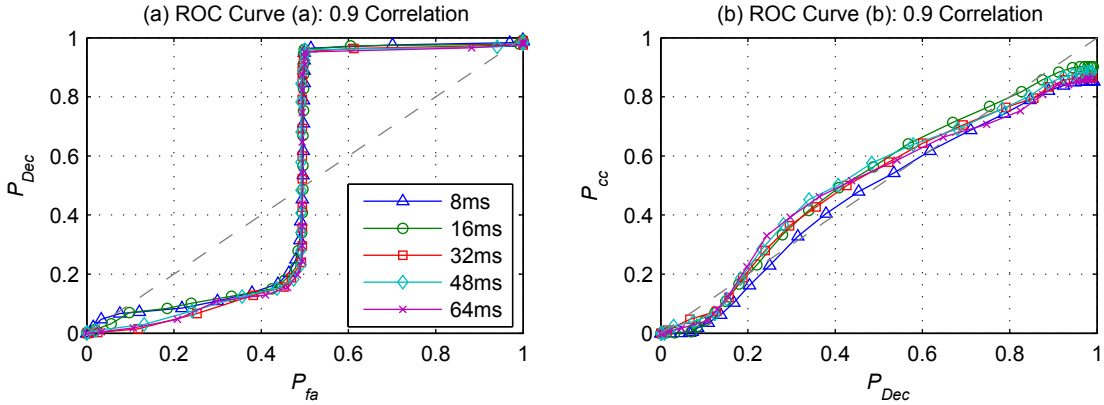


Figure 6.8: The type (a) and type (b) ROC curves for the type II DTW classifier.

regardless of the value of P_{cc} obtained. Despite ignoring P_{cc} during threshold setting it never fell below 0.7, even for single-aspect k-NN, see Fig. 6.9, the worst performing of all the classifiers. Understanding the step in the type (a) ROC curve is helpful in assessing the impact of "unknown" detection on classifiers.

The dramatic step in P_{FA} is a result of one of the unknown targets being significantly more like the known classes than the other. There were two classes of unknown input signal; in the feature space these

are, on average, different distances from the entries of the reference dataset. When the threshold was at a maximum, both unknowns were nearly always incorrectly classified as a known and $P_{\text{FA}} \approx 1.0$. As the threshold dropped the unknown inputs least like the knowns passed the threshold and were correctly declared unknown. Analysis of the raw distances showed the random input to be much further from the known inputs than the simulated propeller; initially, as the threshold fell, only this input class was correctly declared as unknown. Eventually the threshold fell to a level where all the random signals were correctly declared unknown but the all inputs of the propeller-like signal were still misclassified as knowns. At this stage $P_{\text{FA}} = 0.5$. Propeller-like inputs continue to be misclassified as knowns until the threshold had fallen substantially. During the period of lowering the threshold for no improvement in P_{FA} more and more knowns were incorrectly declared unknown and P_{Dec} dropped. It was this drop that lead to the distinctive step.

The final significant feature of the ROC curves is the failure of P_{Dec} to reach 1, as seen in the type (a) curves. The maximum threshold used, as described in section 6.4, relates to the average distance metrics calculated by the classifiers. As for any averaging method some of the distances are much larger than the average value. Analysis of the raw distances showed that in a few situations the distance between a test dataset and reference dataset entry for the same class was much larger than the maximum threshold used. In this situation the test entry will always generate an unknown declaration; since P_{Dec} is the probability of declaring a known class it can never be 1 if some entries are always declared unknown.

6.4.4 Classifier Performance

Figures 6.9 to 6.11 and Table 6.4 show the variation of the evaluation parameters (P_{cc}, RL, P_{FA}, P_{Dec} and P_{Gen}) of section 3.3 across the tests conducted. Table 6.4 provides a "single glance" comparison of the different techniques employed by documenting the parameters for the frame-duration correlation-level pairing deemed to be optimal in the ROC curve analysis. Conversely, the intensity plots of figures 6.9 to 6.11 provide a detailed view. The variation of each evaluation parameter with frame duration and correlation level can be observed providing insight into why the ROC analysis chose a particular pairing as optimal.

It is apparent from Table 6.4 that both type 0 and type II DTW outperform k-NN obtaining a preferable set of evaluation parameters for both the single- and two-aspect reference datasets. The difference was most pronounced in the single-aspect case. Although it must be noted that the k-NN classifier did have a slightly lower false alarm rate, type 0 DTW obtained a P_{cc} of 0.80 and type II DTW 0.90 while k-NN only managed 0.74. The results for RL and P_{Dec} are also preferable for DTW compared to k-NN. In the two-aspect case the difference between the classifiers was smaller, but DTW still performed best. The improved performance of DTW was attributed to its ability to warp the signals for better feature alignment.

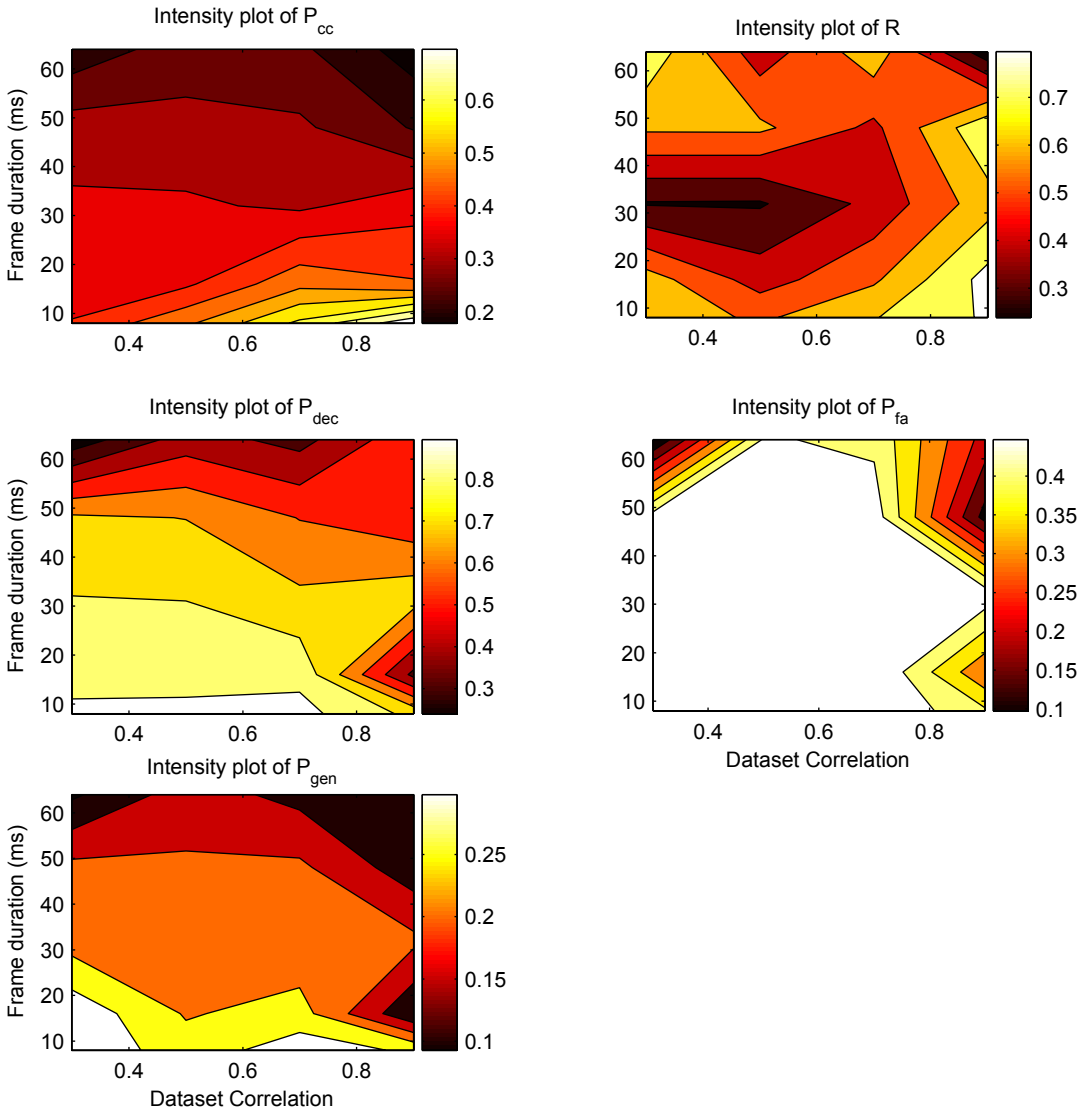


Figure 6.9: The variation of the evaluations parameters with frame duration and dataset correlation, once the "unknown" threshold has been selected, for the k-NN classifier operating with the single-aspect dataset.

In each frame, whether it belongs to the reference or test dataset, the starting phase of the signal is effectively random, see section 4.1. The impact of the randomness on k-NN is shown in Fig. 6.9. The values of P_{cc} and P_{Dec} are higher at shorter frame durations conflicting with the concept of longer frames having more information as reported in (Jahangir et al., 2003b,a). The Euclidean distance between the two distances will be large since the distance between individual samples is increased by the misalignment of the features. These increased distances mean that inputs will appear a long way from similar entries in the reference dataset and are more likely to cross the "unknown" threshold reducing the value of P_{Dec}. The probability of making correct classifications, P_{cc}, is also reduced at longer frame durations since the increased distances between same class signals are comparable to the distances between signals of different class.

In contrast to k-NN DTW coped well with the random starting phase due to its ability to warp the signal and align the features during comparison. Table 6.4 shows that with the single-aspect dataset type II DTW performed the best, while for the two-aspect dataset there was little between the two methods. Type II DTW's success with the single-aspect dataset may be attributed to the limitations the type II local constraints place upon the warping path. In the type 0 case every sample of the input and reference signals must be used; warping is achieved by duplicating individual samples to match to any number of contiguous samples in the reference signal. In the type II case the warping is achieved by allowing individual samples to be skipped over i.e. feature alignment is achieved by deleting single samples. The former method permits many more possible warpings in each comparison and results is dissimilar signals having a low normalized global distance as a result of excessive warping. These low normalized global distances then cause an incorrect declaration to be made by the classifier. In the two-aspect dataset there were more entries in the reference dataset so there was more chance of a good match between an input and a reference. It is believed that this increase in chance of a good match resulted in type 0 DTW's performance equalling that of type II. However, as shall be discussed, the increase of information in the reference dataset also had drawbacks resulting in the overall drop in performance for both DTW methods.

Fig. 6.10 details the evaluation parameters' variations for the type 0 DTW classifier operating on the pure decorrelation time datasets. Both P_{cc} and RL vary in the same manner: improved performance can be obtained by either increasing the frame duration or dataset correlation, although dataset correlation has most impact. P_{FA} was nearly constant across the dataset correlation, but was higher for shorter frame durations. Conversely P_{Gen} showed limited variation with frame duration, but it did improve as the dataset correlation increased. These results were common for all type 0 DTW tests. Although not presented here, the variation of P_{cc}, RL and P_{FA} was the same for the type II DTW when used with a pure decorrelation time dataset.

As shown in Table 6.2 higher dataset correlation results in a larger reference dataset giving more examples per class: the more examples a template based classifier has the better it is expected to perform. The rise in accuracy of the DTW classifier with dataset correlation was not so dramatic when single- and two-aspect reference datasets were used. These datasets had the number of entries per class increased beyond that required by pure correlation distance to ensure they would operate with k-NN. This effectively raised the correlation level of the datasets so DTW was able to perform well even at "low" correlation levels. The improvement with frame duration was attributed to longer frames having higher information content (Jahangir et al., 2003b,a).

An in depth view of the variation of the evaluation parameters with a type II DTW classifier is shown in Fig. 6.11; the classifier is operating on the single-aspect datasets. P_{cc} and RL appear to show substantial variation, but it must be recognized that the intensity scale is small (c.f. the pure decorrelation

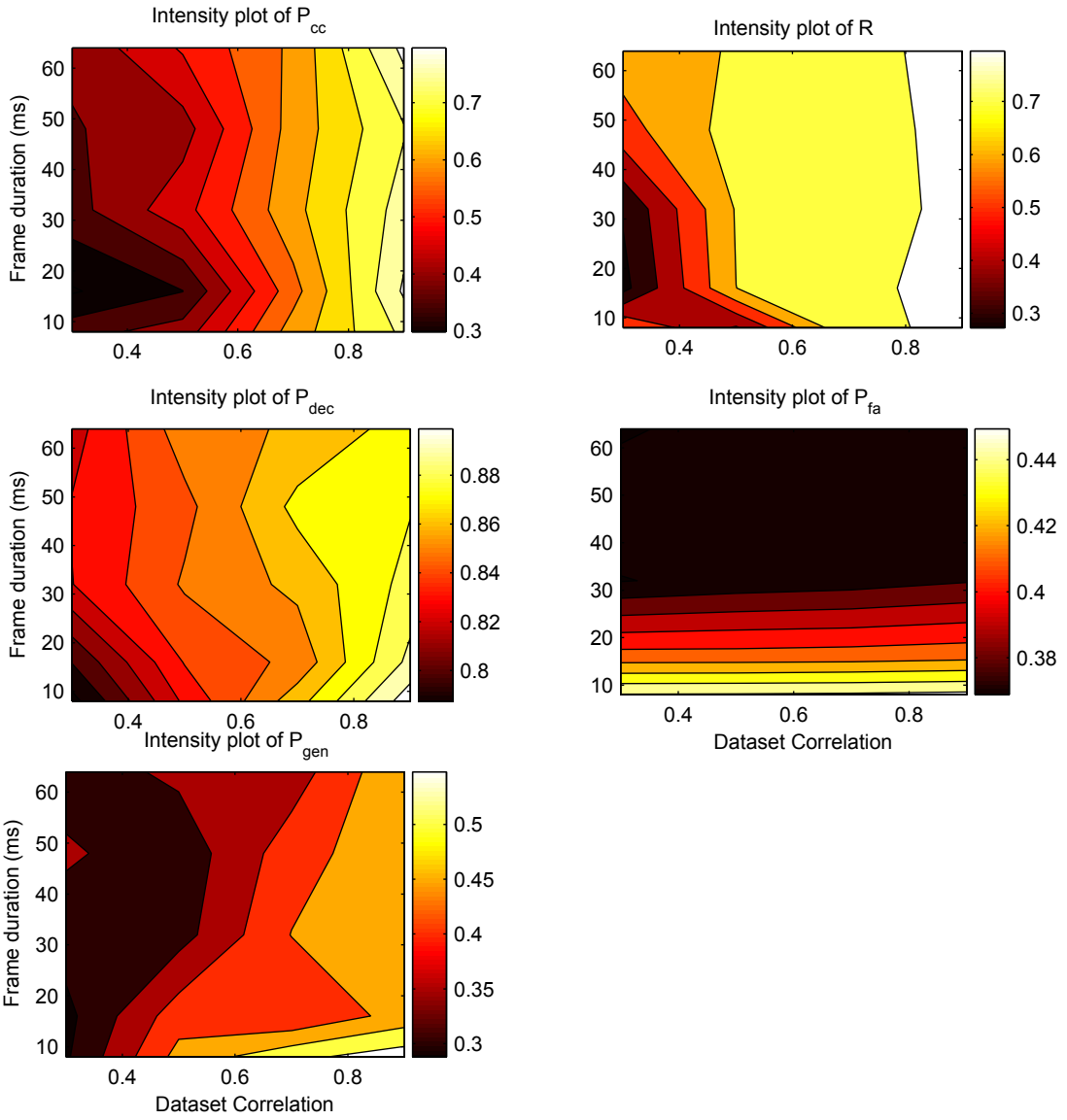


Figure 6.10: The variation of the evaluations parameters with frame duration and dataset correlation, once the "unknown" threshold has been selected, for the type 0 DTW classifier operating with the pure decorrelation time dataset.

time case of Fig. 6.10) so noise on the results is more visible. The P_{FA} plot has a very small intensity range with $P_{\text{FA}} \approx 0.5$ at all frame-duration dataset-correlation combinations. As described above it was often necessary to set the threshold based on the step in P_{FA} that occurred at 0.5, and it is therefore unsurprising that P_{FA} remained at this value.

Some variation in P_{FA} is seen for the type 0 DTW classifier operating with the pure decorrelation time datasets (Fig. 6.10), P_{FA} is larger for short frame durations, and this is supported by the limited variation observed in Fig. 6.11. In Table 6.3 it is seen that for the pure decorrelation time cases the threshold was set using the longest frame duration, 64ms. The threshold is applied to the normalized global distance that, despite normalization, is larger for longer frames since the normalizing constant

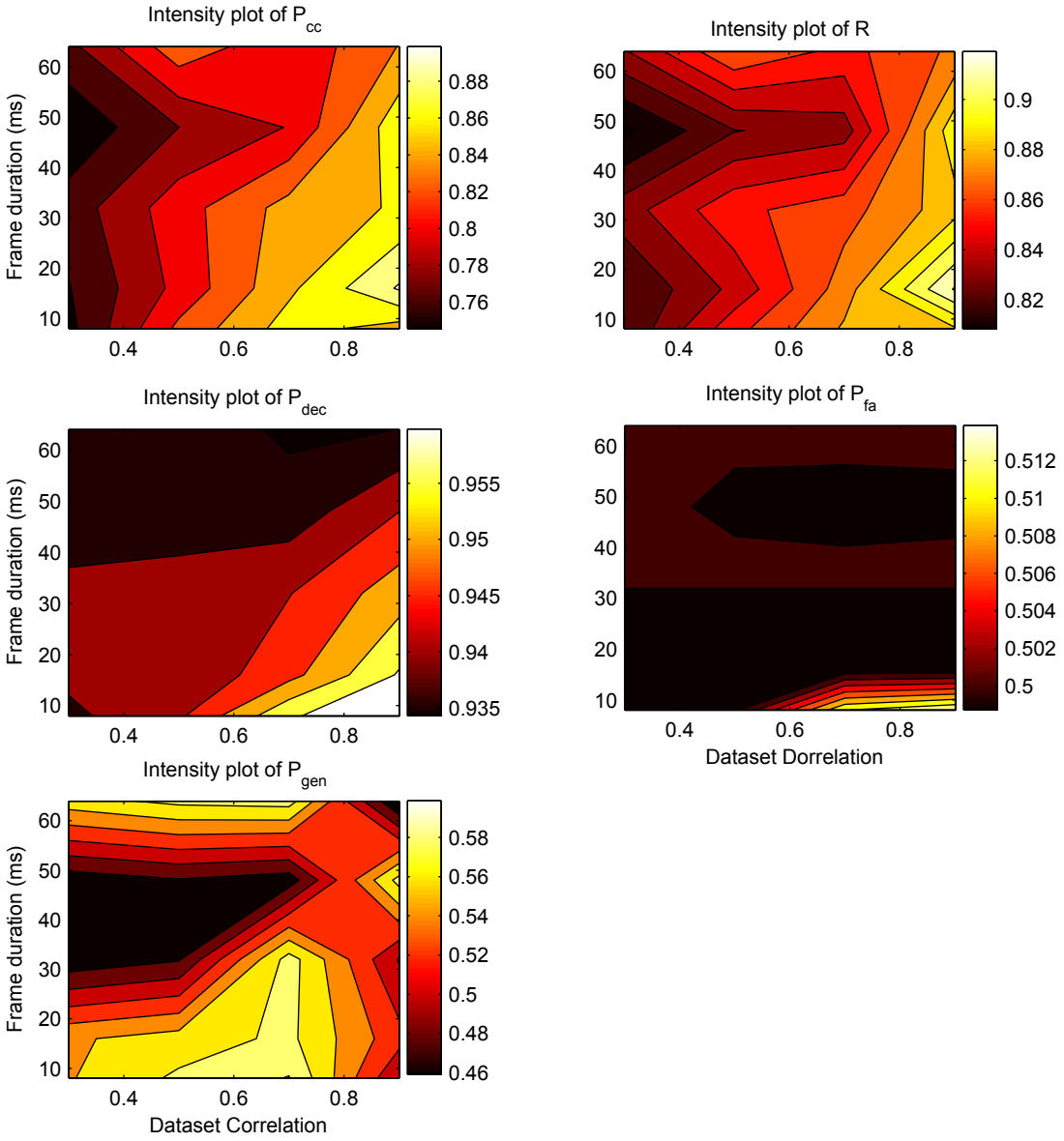


Figure 6.11: The variation of the evaluations parameters with frame duration and dataset correlation, once the "unknown" threshold has been selected, for the type II DTW classifier operating with the single-aspect datasets.

used is an approximation to reduce computation complexity (Myers et al., 1980). By setting the threshold based on long frames it is less likely that the short frames will result in normalized global distances that pass the threshold, so the chance of a false alarm is greater.

Neither k-NN nor DTW were particularly good at generalizing from the reference dataset. For both the single- and two-aspect reference datasets k-NN had $P_{\text{Gen}} = 0.29$ while both type 0 and type II DTW classifiers managed to achieve $P_{\text{Gen}} \approx 0.5$ for all of the dataset types. In the three class situation being considered a classifier that made a random guess would be expected to obtain $P_{\text{Gen}} = 0.33$. k-NN was therefore slightly worse than random at generalizing and DTW only a little better. An analysis of the entries in the confusion matrices suggested the key difficulty for the classifiers in generalizing was

distinguishing the vehicle classes at 90° to radar LOS from personnel. At these angles the main Doppler shift of the vehicles was very small, if not 0 Hz. This makes them appear similar to the personnel reference dataset entries that also have small Doppler shifts since people can not move that quickly. Such confusion suggests that the classifiers were using the MDL frequency in addition to the μ-DS to perform recognition. This operation is to some extent desirable since it is in line with the intuitive steps a human operator may take: any target with a frequency relating to a speed of more than, say, eight miles per hour is unlikely to be personnel since it is difficult for people to maintain this pace for a long time

The performance of the DTW classifiers is similar to the linear discriminant technique where $P_{\text{cc}} = 0.84$ and $RL = 0.85$ when tested under laboratory conditions (Stove and Sykes, 2003). The classifier described in (Stove and Sykes, 2003) does not have the ability to declare unknown nor are any variant classes used, so P_{FA}, P_{Dec} and P_{Gen} are not defined. However, the authors do make reference to how the classifier performance was different when used in the laboratory compared to use in the field. To ensure that the high P_{cc} that was obtained during laboratory testing was maintained for real-world operating scenarios an algorithm that compared the classifications between consecutive scans was employed, although details of this algorithm have not been published. Combining such a technique with DTW would form a valuable extension to the work described here. Also with comparable performance to DTW was the classifier of (Jahangir et al., 2003b,a) that had $P_{\text{cc}} = 0.87$ and $RL = 0.87$, again the other parameters were not defined in this study.

The greedy Gaussian Mixture Model (GMM) classifier described in (Bilik et al., 2006) obtains better results than even the DTW classifiers with $P_{\text{cc}} = 0.98$ and $RL = 0.97$. Again there are no results to give an indication of P_{FA}, P_{Dec} and P_{Gen}. These results are particularly impressive since the greedy GMM classifier had to choose between seven classes, although three of them related to personnel. Interestingly (Bilik et al., 2006) also reports on results obtained for human operators. Their performance is worse than even the k-NN classifier used in this research. They obtain $P_{\text{cc}} = 0.40$, $RL = 0.43$ and $P_{\text{Dec}} = 0.89$. There is no metric for P_{Gen} since no variant targets are used nor P_{FA} as although the operator can declare unknown, allowing P_{Dec} to be defined, no unknown targets are input. It should be noted that the human operators had to choose between six classes, although again three of them related to personnel and the wheeled and tracked vehicle classes were merged into a single vehicle class. Also significant is that the human operators are allowed to listen to four seconds of data, much longer than the maximum 64ms permitted for the DTW classifiers in this study.

Table 6.5 shows the confusion matrices for type II DTW operating with the single- and two-aspect reference datasets side by side. Before discussing the results it is important to understand how to interpret this table. The two-aspect reference dataset contained examples at 60° and 150° for the vehicles and 45° and 135° for personnel, but in the single-aspect case all examples were at 0°. This difference effects the entries used to calculate the evaluation parameters. For example, the probability of correctly classifying

a wheeled vehicle in the single-aspect case is the probability that 'Wheeled Vehicle 0°' was declared 'Wheeled Vehicle', while for the two-aspect case it is the average of the probabilities that 'Wheeled Vehicle 30°' and 'Wheeled Vehicle 150°' were declared 'Wheeled Vehicle'. This concept was then extended to calculate all the evaluation parameters from section 3.3.

The table indicates that superior recognition was achieved with the single-aspect datasets for vehicle classes, but that the personnel class was better identified in the two-aspect case. Overall the increase in confusion between the two vehicle classes resulted in the two-aspect datasets having worse performance than the single-aspect. This was a surprising result since the two-aspect reference datasets contained more examples of the reference classes. Observing the change of the μ-DS with aspect angle in figures 4.8 to 4.13 can help explain the change in performance: when the wheeled vehicle is head-on to the radar LOS its moving components are occluded and the μ-DS shows no significant sidebands; at 30° to head-on, however, sidebands do appear. With the sidebands present there is much more similarity between the wheeled and tracked vehicle spectrograms and the increased confusion when two-aspects are used was attributed to this similarity. Furthermore, an examination of personnel spectrograms showed them to be almost identical from all aspect angles: the features of the spectrogram were consistent whether viewed from 0°, 45°, 135° or 180°. Adding an extra aspect to the reference dataset was therefore comparable to increasing the dataset correlation for personnel accounting for the increase in personnel classification performance in the two-aspect case.

6.4.5 Limitations Of The Classifiers

Both k-NN and DTW are template based classifiers, and as such they share a common limitation: the reference dataset must be present to perform a classification. As has been discussed, the best performance tends to occur when there is a high dataset correlation; Table 6.2 documents how high correlations also result in large datasets. There are two problems with large datasets: each input must be compared with every entry of the reference dataset so when the dataset is large classification will take longer; and the larger the dataset the more memory is required to store it. It is possible to overcome the former of these difficulties, but not the latter.

Although classification time was not formally recorded in this study it has been shown that k-NN is able to perform much faster than DTW (Smith et al., 2006b). Initially the DTW algorithm was implemented in Matlab, but ran so slowly that its use with large reference datasets was not feasible. Recoding the algorithm in C++ provided an implementation that could classify with the largest reference dataset in under 10s. k-NN on the other hand completed classifications in under 1s for even the largest dataset. Further optimization of the DTW algorithm is possible and should enable it to operate nearly as quickly as k-NN (Keogh and Ratanamahatana, 2002) and see Appendix B.

Unfortunately there is no method for reducing the size of the reference dataset while keeping its

information content the same. In this study only three classes of target are used; future systems may require the detection of further classes leading to even larger reference datasets. Under such situations it may be necessary to use a technique such as Principal Component Analysis (PCA) to reduce the size of each entry in the dataset, although this would necessitate a move from the time domain into a new domain of reduced dimension. An alternative would be to move to a learning classification algorithm, such as Artificial Neural Networks (ANNs) or Hidden Markov Models (HMMs), where the reference dataset is used offline, as training data, and is not needed for actual classification.

6.5 Summary

It is possible to conduct μ-DS based radar ATR using minimal feature extraction by working in the time domain. Rather than employ the detailed process described in Chapter 5 the classification in this section was based on feature vectors comprised of nothing more sophisticated than the power at each sample of the data frame. Classification of this data was performed using the DTW and k-NN algorithms that were introduced at the start of the chapter.

The first part of the time domain investigation studied use of the signal power compared to I- and Q-values in the feature vectors. These tests used a simplified version of the DTW classifier that did not include an "unknown" detection threshold. It was observed that the sample-power feature vectors gave results comparable to I- and Q-value cases. After this investigation, the k-NN classifier was included in the testing and the "unknown" threshold restored to the DTW classifier.

Prior to main classification testing, the dataset sizes were reviewed and the level of the "unknown" threshold set. It was observed that higher dataset correlation levels and the shorter the frame durations resulted in larger reference datasets. The setting of the "unknown" threshold for the classifiers demonstrated the use of the ROC curve analysis developed in Chapter 3. The type (a) ROC curves produced had a distinctive shape and this was explained as resulting from there being just two unknown classes in the test dataset.

The main testing included investigated the impact of frame duration, dataset correlation, number of target aspects in the reference dataset and the type of classifier on radar ATR performance. The DTW classifiers were observed to give better performance than the k-NN classifiers. When a single target aspect was included in the reference dataset type II DTW performed better than type 0, but when two aspects were included in the reference dataset there was little between the two implementations. For both types of classifier the system performance improved with dataset correlation level. The result was attributed to high correlation levels resulting in more entries in the reference dataset; template based classifiers tend to perform better with more entries in the reference dataset (Duda et al., 2005). With regard to frame duration: the two types of DTW showed improved performance with longer frames, a result consistent with the published literature. However, k-NN's performance fell as frame duration

increased. The result was attributed to feature misalignment between the frames being more significant at longer durations. Unlike DTW the k-NN algorithm does not compensate for misalignment of the features so the classifier's performance fell as the frames became longer. It was noted that none of the classifiers performed particularly well when it came to recognizing the variant classes of the test dataset. The DTW performance was better than a random selection of declaration class. This suggesting it was capable of performing a limited degree of generalization in the time domain.

It was decided that following the time domain testing, investigation should switch to frequency domain. The problem of random starting phase of the μ-DS was expected to be mitigated in this domain and it was anticipated that the k-NN classifier would show improved performance alongside improvement in the generalization capability of all the classifiers.

Table 6.2: Sizes of the reference datasets used in this study.

	Number of entries in the reference dataset for the:														
	Single-aspect case					Two-aspect case					Pure decorrelation time case				
Correlation	8ms	16ms	32ms	48ms	64ms	8ms	16ms	32ms	48ms	64ms	8ms	16ms	32ms	48ms	64ms
0.3	69	69	70	72	70	138	138	140	143	145	6	9	11	15	16
0.5	106	88	86	85	92	1125	484	232	189	182	21	22	29	31	35
0.7	421	228	140	109	100	734	520	335	249	205	68	60	72	79	66
0.9	536	334	185	172	129	1470	762	431	321	245	423	233	143	116	96

Table 6.3: The minimum, maximum and selected values of the thresholds for all dataset types. The dataset correlation and frame duration on which the threshold setting was based are shown.

Test	Threshold Range Min	Max	Threshold Selected	Dataset Correlation	Frame Duration (ms)
1 Aspect k-NN	1×10^4	5×10^8	$3.57 \times 10^{6\dagger}$	0.9	8
1 Aspect Type 0 DTW	6×10^5	6×10^{11}	$5.63 \times 10^{9\ \dagger}$	0.9	16
1 Aspect Type II DTW	6×10^5	6×10^{11}	$5.54 \times 10^{10\dagger}$	0.9	16
2 Aspect k-NN	9×10^3	3×10^7	5.60×10^6	0.5	16
2 Aspect Type 0 DTW	3×10^5	8×10^{11}	8.12×10^9	0.9	48
2 Aspect Type II DTW	3×10^5	8×10^{11}	2.25×10^{10}	0.9	64
Simplified Intensity Type 0 DTW	6×10^5	6×10^{11}	$1.60 \times 10^{10\dagger}$	0.9	64
Simplified Intensity Type II DTW	6×10^5	6×10^{11}	8.92×10^{10}	0.9	64

† Linear interpolation was used to find this threshold value.

Table 6.4: The classifier evaluation parameters for all permutations of dataset and classifier type. The values given are for the frame-duration dataset-correlation pairing and unknown threshold listed in Table 6.3.

	Single Aspect Dataset			Two Aspects Dataset			Simplified Intensity	
Parameter	k-NN	Type 0 DTW	Type II DTW	k-NN	Type 0 DTW	Type II DTW	Type 0 DTW	Type II DTW
P_{cc}	0.74	0.80	0.90	0.71	0.75	0.76	0.81	0.84
RL	0.81	0.94	0.92	0.74	0.77	0.77	0.87	0.88
P_{Gen}	0.29	0.47	0.53	0.29	0.53	0.41	0.50	0.49
P_{FA}	0.41	0.39	0.50	0.52	0.50	0.50	0.38	0.50
P_{Dec}	0.77	0.81	0.96	0.95	0.96	0.94	0.87	0.95

Table 6.5: A comparison of type II DTW for the single- and two-aspect datasets. The table is a confusion matrix in which two probabilities are given for each declared class: the first for the single-aspect dataset; the second for the two-aspect dataset.

	Declared Class							
	Wheeled Vehicle		Tracked Vehicle		Personnel		Unknown	
Input Class	Single	Two	Single	Two	Single	Two	Single	Two
Wheeled Vehicle 0°	0.98	0.24	0.00	0.23	0.02	0.48	0.00	0.05
Wheeled Vehicle 30°	0.40	0.79	0.41	0.10	0.19	0.10	0.00	0.00
Wheeled Vehicle 60°	0.16	0.63	0.74	0.28	0.10	0.09	0.00	0.00
Wheeled Vehicle 90°	0.50	0.09	0.04	0.01	0.46	0.90	0.00	0.00
Wheeled Vehicle 120°	0.23	0.46	0.69	0.43	0.06	0.10	0.02	0.00
Wheeled Vehicle 150°	0.30	0.78	0.64	0.19	0.06	0.03	0.00	0.00
Wheeled Vehicle 180°	0.85	0.14	0.00	0.19	0.00	0.21	0.15	0.47
Tracked Vehicle 0°	0.09	0.44	0.90	0.55	0.00	0.00	0.01	0.01
Tracked Vehicle 30°	0.26	0.29	0.71	0.67	0.03	0.03	0.00	0.00
Tracked Vehicle 60°	0.17	0.75	0.79	0.22	0.04	0.01	0.01	0.02
Tracked Vehicle 90°	0.24	0.34	0.28	0.12	0.42	0.49	0.06	0.06
Tracked Vehicle 120°	0.17	0.71	0.76	0.23	0.03	0.01	0.04	0.05
Tracked Vehicle 150°	0.18	0.65	0.79	0.33	0.04	0.02	0.00	0.00
Tracked Vehicle 180°	0.15	0.17	0.43	0.34	0.00	0.01	0.42	0.47
Personnel 0°	0.17	0.06	0.00	0.01	0.83	0.92	0.00	0.01
Personnel 45°	0.11	0.03	0.00	0.00	0.89	0.97	0.00	0.00
Personnel 135°	0.24	0.00	0.00	0.00	0.76	1.00	0.00	0.00
Personnel 180°	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Simulated Propeller	0.89	0.01	0.00	0.00	0.11	0.99	0.00	0.00
Random Signal	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00

Chapter 7

Monoperspective Frequency Domain Classification

In this chapter testing in the frequency domain is considered. The effects of forming the feature vectors through frequency normalization and application of Principal Component Analysis (PCA) are discussed. The naïve Bayesian classifier is introduced, alongside how frequency spectrum work affects Dynamic Time Warping (DTW) and k-Nearest Neighbour (k-NN). Results from the frequency domain are then presented and discussed. The discussion includes comparison, where appropriate, with the results of the performance prediction.

7.1 Modification To Dynamic Time Warping and k-Nearest Neighbour

It was decided to move from the time domain to the frequency domain after it was realized that the k-Nearest Neighbour (k-NN) classifier was struggling to cope with the effect of phase on the Micro-Doppler Signature (μ-DS). The Dynamic Time Warping (DTW) and k-NN classifiers described in sections 6.1 and 6.2 required no modification to work in the frequency domain; it was only necessary to adjust the feature extraction process. In the time domain approach the power samples of each frame of the received radar signal were treated as elements of a feature vector. Moving to the frequency domain does nothing to change the structure of the data, it is only the *meaning* of the data that has changed. Where as before successive elements of the feature vector represented the change in the signal power with time, in the frequency domain successive elements represent the power at successive frequencies. The classifiers are not concerned with such semantics and whether operating in the time or frequency domains they are still processing vectors with real valued elements. Indeed the formal definition of the DTW algorithm does not require the warping axis to be time, only some monotonically increasing variable such as the frequency of a power spectrum (Sankoff and Kruskal, 1999). The only changes to these

classifiers, therefore, were in the preparation of the data. The frequency domain was also suitable for application of a further classifier, the naïve Bayesian, that uses the reference dataset during training only and does not require its presence when classifying

The process by which the frequency domain data was prepared was described in section 5.4. For DTW and k-NN no reduction of the feature vector dimension was performed and the amplitude scale of the power spectra was left as linear rather than the Decibel (dB) scale used during Principal Component Analysis (PCA). This processing was applied to both the reference and test dataset ahead of using the classifiers. In the real system, however, these steps would be the feature extraction of Fig. 3.1 and would be performed during operation rather than in advance.

7.2 Naïve Bayesian Classification

The naïve Bayesian classifier is based around Bayes's Theorem and makes the naïve assumption that the elements of an input feature vector are statistically independent. Despite this simplifying assumption the classifier achieves good results even when intricate multivariate datasets are used (Hand and Yu, 2001; Zhang, 2004). If the elements of a feature vector $\mathbf{x}$ are statistically independent then:

$$\mathrm{P}(\mathbf{x}) = \prod_{i=1}^{D} \mathrm{P}(x_i) \tag{7.1}$$

where the set $\{x_i\}$ are the elements of the vector $\mathbf{x}$ and D is the number of elements. If there is a set C of classes defined:

$$C = \{c_1, c_2, \ldots c_m\} = \{c_j\} \text{ where } \{j \in \mathbb{N} : j \leq N_c\}, \tag{7.2}$$

where N_c is the number of classes, then the probability of a particular class, based on a given feature vector and the prior probability of seeing the class, $\mathrm{P}(c_j)$, is:

$$\mathrm{P}(c_i|\mathbf{x}) \propto \mathrm{P}(c_j) \prod_{i=1}^{D} \mathrm{P}(x_i|c_j), \tag{7.3}$$

From (7.3) the following classification rule can be developed (Duda et al., 2005):

$$c = \underset{c_j \in C}{\operatorname{argmax}}\, \mathrm{P}(c_j|\mathbf{x}) \propto \underset{c_j \in C}{\operatorname{argmax}}\, \mathrm{P}(c_j) \prod_{i=1}^{D} \mathrm{P}(x_i|c_j). \tag{7.4}$$

That is, select the class c to maximize the likelihood $\mathrm{P}(c_i|\mathbf{x})$, which is proportional to $\mathrm{P}(c_j) \prod_{i=1}^{D} \mathrm{P}(x_i|c_j)$. (7.4) is the Bayes's decision rule and minimizes the probability of error. To complete the classifier the likelihoods $\mathrm{P}(x_i|c_j)$ are required.

The simplest approach to selecting $\mathrm{P}(x_i|c_j)$ is to assume that all of the probabilities have a normal distribution. By definition the likelihoods $\mathrm{P}(x_i|c_j)$ are in the interval [0,1]. As the feature vector

dimensionality, D, increases then $\lim_{D\to\infty}\prod_{i=1}^{D}\mathrm{P}(x_i|c_j) = 0$. $\mathrm{P}(c_j|\mathbf{x})$, from (7.3), is therefore very small, and difficult to store reliably in digital memory, for large D. This can be overcome by using the log likelihood, i.e. taking the log of the normal distribution, leading to (Duda et al., 2005):

$$\mathrm{L}(x_i|c_j) = \ln\left(\mathrm{P}(x_i|c_j)\right) = -\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2} - \ln\sqrt{\sigma_{ij}^2 2\pi} \tag{7.5}$$

and

$$\mathrm{L}(c_i|\mathbf{x}) \propto \sum_{i=1}^{n}\mathrm{L}(x_i|c_j) + \ln\left[\mathrm{P}(c_j)\right], \tag{7.6}$$

where μ_{ij} and σ_{ij}^2 are the mean and variance for the i^{th} element of the class c_j. (7.4) then becomes

$$c = \underset{c_j\in C}{\operatorname{argmax}}\,\mathrm{L}(c_j|\mathbf{x}) \propto \underset{c_j\in C}{\operatorname{argmax}}\sum_{i=1}^{n}\mathrm{L}(x_i|c_j) - \ln\left(\mathrm{P}(c_j)\right). \tag{7.7}$$

Detection of unknown classes, those outside the set C can be performed by applying the "unknown" threshold from section 3.1 to the internal scores of the classifier (Smith et al., 2008a). In the naïve Bayes classifier the scores are the log likelihoods and the threshold will need to ensure that the value of $\mathrm{L}(c_j|\mathbf{x})$ from (7.7) is large enough that the input can be considered as being a member of C.

7.3 Testing Strategy

There were several objectives for the frequency domain μ-DS classification tests:

1. Investigate how removing the dependence of feature vectors on signature starting phase effected performance.
2. Perform a more comprehensive feature extraction process to try to improve the separability of the classes.
3. Attempt to predict the performance of a classifier before implementing it.
4. Allow the use of the naïve Bayesian classifier to compare its performance with DTW and k-NN.

To meet these objectives a series of test datasets were required.

Although the frequency domain pre-processing required that the entire dataset creation process be repeated, it was decided to keep the datasets as similar as possible to those used during the time domain testing in Chapter 6. Therefore the same frame durations, 8 ms, 16 ms, 32 ms, 48 ms and 64 ms, and dataset correlations, 0.3, 0.5, 0.7 and 0.9, were used. For the DTW and k-NN classifiers two dataset were prepared while many were used for the naïve Bayesian classifier.

Just as for the time domain case single and two aspect datasets were produced. Again, in the single aspect datasets only examples of the targets travelling head-on (at 0°) to the radar Line Of Sight (LOS) were included in the reference dataset; and in the two aspect reference datasets the vehicle targets were represented with examples at 30° and 150°to the radar LOS while the personnel target was represented at

Table 7.1: Summary of datasets for Frequency domain classification.

Name	Aspects In Reference Dataset	Aspects In Test Dataset
Template Single Aspect Dataset	Head-on (0° to LOS)	All available
Template Two Aspects Dataset	30° and 150° to LOS for vehicles 45° and 135° to LOS for personnel	All
Bayes Single Aspect $x\%^*$ Dataset	Head-on	All available
Bayes Two Aspects $x\%^*$ Dataset	30° and 150° to LOS for vehicles 45° and 135° to LOS for personnel	All
Bayes All Aspects $x\%^*$ Dataset	All available	All available

* Where $x\%$ is the percentage variance remaining in the data after PCA and may range from 45% to 95% in steps of 5%.

45° and 135°. As described in section 7.1 the datasets to be used with the template matching classifiers (DTW and k-NN) did not have dimensionality reduction performed on them and had linear amplitude scales. These datasets will be referred to as the *template single aspect dataset* and *template two aspects dataset*.

The basic concept behind the naïve Bayesian classifier datasets was the same as the template classifier ones: have single and two aspect datasets. During testing it was realized that there was a strong connection between the number of aspects in the reference dataset and the classifier performance so a third case was considered in which examples from all aspect angles were included in the reference dataset. Unfortunately, this dataset was unable to test the generalization properties of the classifier, since there were no aspects in the tests dataset that could be considered variants. These datasets will be referred to as the *Bayes single aspect dataset*, *Bayes two aspects* and *Bayes all aspects* datasets respectively. However, there is a further stage in the creation of the Bayes datasets to consider that relates to the dimension reduction.

The amount of variance retained in the feature vector after PCA had been applied was expected to effect classifier performance. As such each of the Bayes datasets was created several times retaining a different percentage of the total variance each time. Percentages ranging from 45% to 95%, in steps of 5%, were used. The changes in variance had no effect on the number of entries in the dataset nor the correlation time used to select which frames entered the reference and test datasets. Instead the different percentage variances result in feature vectors of different lengths. To identify the different amount of variance in a dataset the percentage was appended to dataset name. For example the two aspects dataset with 70% variance was the *Bayes two aspects 70%* dataset. The details of the datasets are summarized in Table 7.1.

With the datasets created the testing strategy differed slightly between the classifiers. For the template based systems (DTW and k-NN) the tests were conducted in the same manner as for the time domain, see section 6.3, with the "unknown" threshold level being selected using Receiver Operating Characteristic (ROC) curve analysis followed by analysis of the classifiers using this threshold. The

Test	Threshold Range Min	Threshold Range Max	Threshold Selected	Dataset Correlation	Frame Duration (ms)
1 Aspect k-NN	2.6×10^4	2.2×10^{10}	1.2×10^8	0.5	8
1 Aspect Type 0 DTW	3.0×10^6	8.1×10^{16}	1.1×10^{14}	0.9	8
1 Aspect Type II DTW	1.4×10^6	5.7×10^{16}	1.1×10^{16}	0.9	64
2 Aspects k-NN	2.9×10^5	7.1×10^9	1.6×10^8	0.9	8
2 Aspects Type 0 DTW	5.6×10^7	4.3×10^{16}	7.5×10^{13}	0.9	8
2 Aspects Type II DTW	3.1×10^7	7.8×10^{16}	9.5×10^{13}	0.9	8

Table 7.2: The minimum, maximum and selected values of the thresholds for all template-classifier frequency datasets. The dataset correlation and frame duration on which the threshold setting was selected is also shown.

naïve Bayesian classifier, however, required a slightly different approach.

Before the actually classification testing could begin with the naïve Bayesian it was necessary to select the percentage variance that would be kept in the feature vectors. This selection was made by disabling the "unknown" threshold, ignoring the variant and "unknown" targets in the test dataset, and then running the classifier at each percentage variance. The only evaluation metric to be considered was P_{cc} and the variance that resulted in the highest P_{cc} was then selected for further testing. For the selected percentage variance an estimate of the classifier performance was made using the Bhattacharyya bound on the probability of error. The same testing method, of first setting the "unknown" threshold and then performing the classification tests, as was used for the template classifiers was then applied. It was necessary to use a different level of "unknown" threshold for each frame duration for the naïve Bayesian classifier as the dimension of the feature vector had a marked impact on threshold level required.

The results for the frequency domain testing will be broken down as follows:

- The results for the template based classifiers will be analysed;
- The impact of reducing the feature vector dimension will be considered;
- The naïve Bayesian classifier results will be considered and contrasted with the performance predictions;
- The different classifiers will be compared and their limitations outlined.

7.4 Results And Discussion

7.4.1 Template Based Classification

The feature extraction process used for frequency domain classification did not effect the distribution of data frames between the reference and test datasets. The selection of frames for the reference dataset was performed by the correlation time method described in section 5.3 and the size of the datasets produced was the same as in Table 6.2. The content of each data frame was now the normalized power spectrum of the signal and this lead to different results for the "unknown" threshold level.

The top level results for the setting of the "unknown" threshold are shown in Table 7.2. Alongside

	Single Aspect Dataset			Two Aspect Dataset		
Parameter	k-NN	Type 0 DTW	Type II DTW	k-NN	Type 0 DTW	Type II DTW
P_{cc}	0.65	0.82	0.80	0.73	0.73	0.73
RL	0.78	0.85	0.62	0.76	0.74	0.74
P_{Gen}	0.51	0.57	0.57	0.65	0.66	0.66
P_{FA}	0.50	0.50	0.50	0.50	0.50	0.50
P_{Dec}	0.86	0.94	0.93	0.85	0.92	0.94

Table 7.3: The classifier evaluation parameters for all permutations of dataset and classifier type. The values given are for the frame-duration dataset-correlation pairing and unknown threshold listed in Table 7.2.

the minimum, maximum and selected threshold the dataset correlation and the frame duration on which the selection was based are indicated. The threshold values in Table 7.2 are greater than those for the time domain in Table 6.3. This increase is attributed to the amplitude range of the output of the Discrete Fourier Transform (DFT) process being greater than the amplitude range of the time domain signals. Unlike the time domain case none of the thresholds were set using linear interpolation to obtain $P_{\text{cc}} = 0.8$: in the frequency domain all the type (a) ROC curves had the distinctive step at $P_{\text{FA}} = 0.5$ so the threshold that gave maximum P_{Dec} at $P_{\text{FA}} = 0.5$ was selected. It is interesting to note that in Table 7.2 for all but the single aspect type II DTW case the threshold was selected based on the 8 ms curve indicating that short frames gave better performance. This is in contrast to the time domain case, where 64 ms, gave best performance for DTW classifiers due to their higher information content. The reasons for this change in behaviour will become apparent when the general classifier performances are discussed.

The results for the template classifier testing are summarised in Table 7.3. These results can be compared directly with Table 6.4, that summarizes the performance of the time domain versions of the template classifiers. The comparison shows that in the frequency domain the classifiers are better able to generalize, i.e. the values of P_{Gen} are higher for the frequency domain. Conversely, the other frequency domain evaluations parameters are either similar or worse than their time domain counter parts depending on the number of aspectss in the reference dataset. For a single aspect reference dataset the time domain classifiers perform better while for the two aspects cases the performance is similar in both domains. These results were surprising since it had been hypothesised that operation in the frequency domain, where the starting phase of the signature is irrelevant, would improve the classifiers' performances.

Another even more unexpected result was the variation of the classifier performance with frame duration. Fig. 7.1 shows the variation of the evaluation parameters with both frame duration and dataset correlation level for the type II DTW classifier operating with a two aspects reference dataset. The first observation is that the intensity plot for the probability of false alarm, P_{FA}, indicates 0.5 for all permutations of frame duration and dataset correlation, a result of the way the "unknown" threshold was selected. For all of the remaining parameters there are two clear general trends: that higher dataset

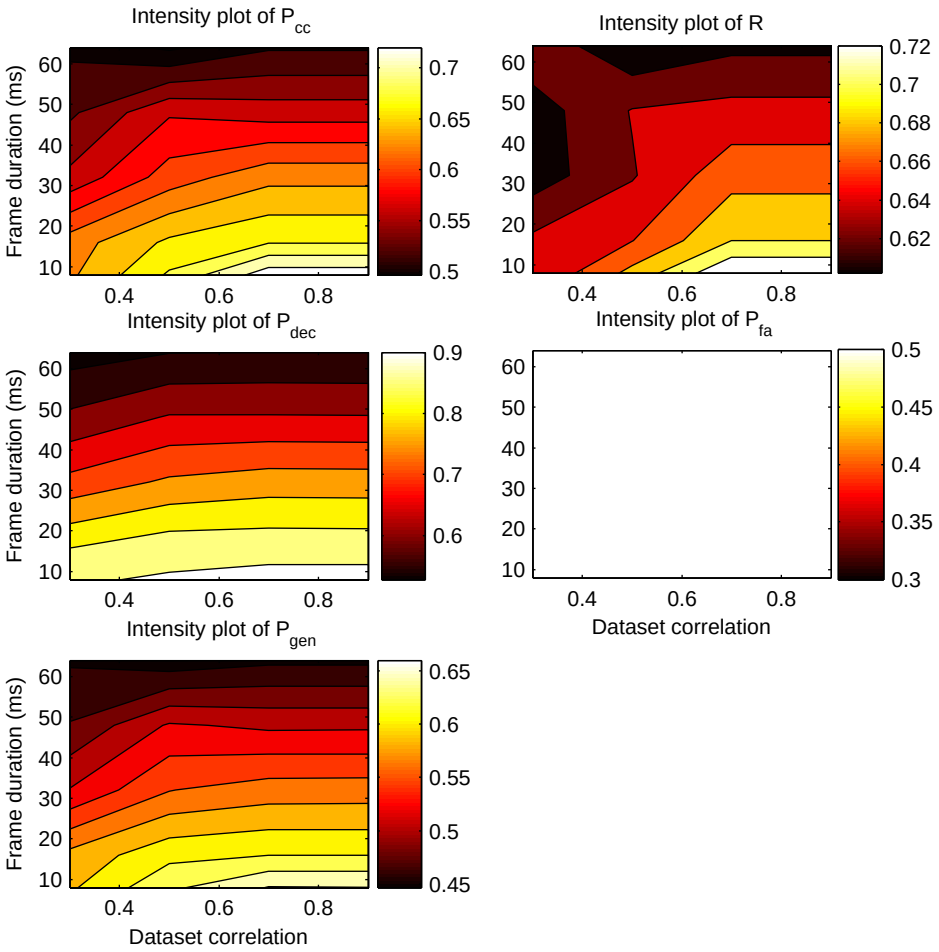


Figure 7.1: Detailed variation of evaluation parameters for the two aspects type II DTW classifier in the frequency domain.

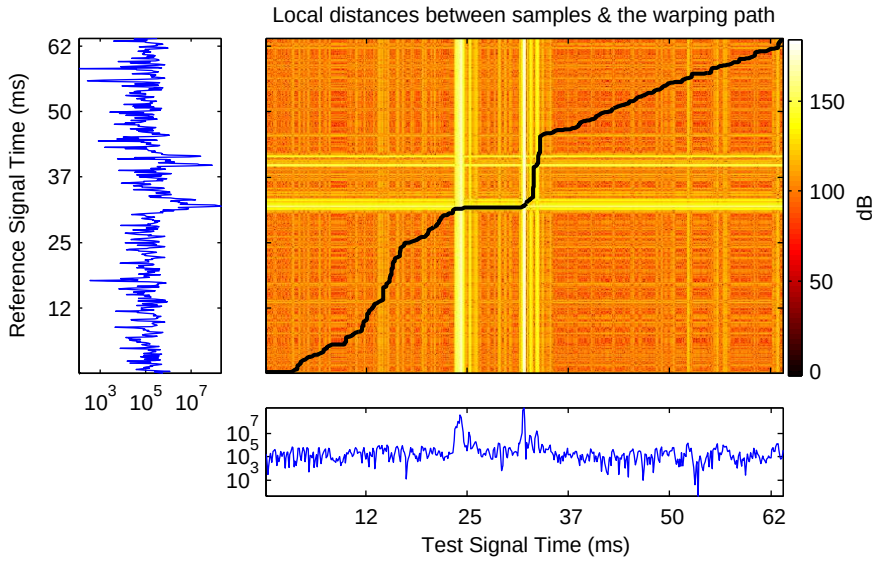


Figure 7.2: Operation of type 0 DTW in the frequency domain.

correlation gives better performance and that short frames give better performance. As was discussed in section 6.4 that higher correlation datasets give better performance is no surprise since there are more examples of each class within them. Conversely, that shorter frames should give better performance is unexpected as longer frames are considered to contain more information about the target (Jahangir et al., 2003b,a). Although a similar result was seen, and explained, in the time domain for the k-NN classifier detailed investigation of the frequency results showed a different mechanism to be at work.

The operation of the type 0 DTW when contrasting a tracked vehicle target input with a tracked vehicle entry of the reference dataset is shown in Fig. 7.2. The frame duration is 64ms, the reference signal is shown to the left of local distance map, the test signal is shown at the bottom and the warping path is superimposed over the map. Due to the frequency normalization process the region of the feature vector that contains the μ-DS is around 32 ms into the test and reference frames. It would seem reasonable therefore to expect the local distances to be at a minimum for times around 32 ms in both the signal since this is where the (hopefully) matching signature is positioned. Instead the figure shows that the local distance is actually close to maximum at these times. Consideration of why this is will be held back for the time being and the impact on the classifier discussed first.

DTW is only able to apply a warping along the time axes of two data series and it does this to minimize the global distance between the presented feature vectors (Sankoff and Kruskal, 1999). The local distance between individual samples, that is used to assess where along the time axis features are, depends on the amplitude of each of the samples and it has been shown that DTW runs into difficulty when there is a large discrepancy between the amplitude of the reference and test feature vectors (Keogh and Pazzani, 2001a). In these situations the algorithm has a tendency to perform unrealistic warpings to align samples of similar amplitude rather than align similar features. This can be seen in Fig. 7.2. The

Frame Duration (ms)	8	16	32	48	64
MDL Standard Deviation	1.4×10^8	6.3×10^8	3.2×10^9	5.5×10^9	1.2×10^{10}
1$^{\text{st}}$ Frequency Bin	2.1×10^4	4.8×10^4	8.9×10^4	1.3×10^5	3.3×10^5

Table 7.4: The standard deviation of the MDL and first frequency bin for different frame durations.

warping path runs approximately diagonally, the ideal path, through the regions of the local map that correspond to noise in the two feature vectors since the noise amplitude is the approximately the same in both vectors. When the signature region of the feature vectors is reached, however, there are large warpings as DTW attempts to align samples of similar amplitude.

As a result of this analysis of the operation of DTW it is clear that the transition to the frequency domain has unintentionally changed the feature vectors to have properties that are incompatible with the DTW process. Furthermore, Fig. 7.2 indicates that large parts of the frequency domain feature vectors are noise. The noise is not affected by the target and so, since noise looks like noise, all the feature vectors, be they reference or test, have large sections that appear similar. This will increase the overall similarity of the vectors compared with the time domain case where the μ-DS is spread throughout the data frame. Although the above paragraphs have described the workings of the frequency domain DTW classifiers they have not yet explained why the shorter frames gave better performance.

The shorter the frame duration the less amplitude variation there was in the μ-DS so the better DTW was able to perform. The standard deviation, calculated from the 0.9 correlation datasets, of the Main Doppler Line (MDL) and first frequency bin, that contained only noise, are contrasted for the different frame durations in Table 7.4. It is apparent that there is much more variation, between four and five orders of magnitude, in the power of the MDL of the μ-DS than there is in the noise and that the amount of variation increases with the frame duration. As was discussed above DTW performs poorly when confronted with large amplitude variations so it is not surprising that shorter frames give better performance in the frequency domain. The reason for the larger variation in μ-DS power at longer frames can be attributed to target scintillation. It is known that the Radar Cross Section (RCS) of a radar target is strongly dependent on aspect angle (Skolnik, 1990) and even a small amount of target motion is enough to cause an aspect change that significantly effects the power of the backscatter signal. Since there is inevitably greater aspect angle change for longer frame durations it is reasonable to expect greater power fluctuations for these frames. Although not studied in detail in this work, some methods that could be employed to limit these effects and improve the DTW performance have been investigated.

The simplest way to counter the large signal power fluctuations that degrade DTW performance would be to use a dB scale rather than a linear one. Such scales are common place in radar precisely because of the large amplitude fluctuations that a signal may exhibit. Additionally it is possible to modify the DTW algorithm to consider the rate of change of local distance between samples as well as the distance itself and this has been shown to improve its performance in situations where there is a

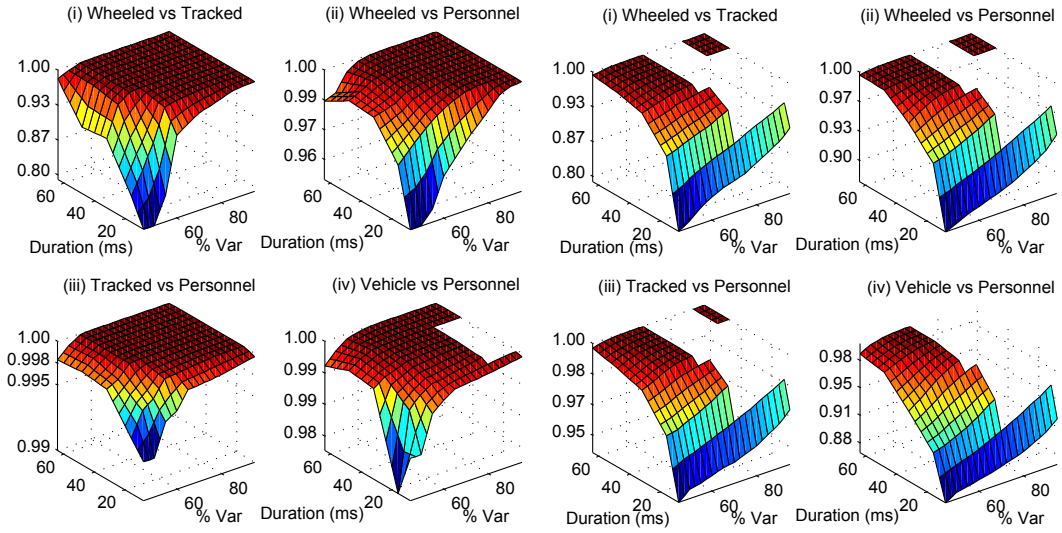
large amplitude difference between the signals being compared (Keogh and Pazzani, 2001a). Lastly, one of the first effects of dimension reduction by PCA is the removal of the noise elements from the feature vectors. Had such processing been applied to the feature vectors here the problem of having large regions of each vector exhibiting similar features regardless of target class could have been removed. In order to keep the investigation moving forward it was decided to investigate the impact of these techniques on the naïve Bayesian classifier rather than the DTW method.

7.4.2 Estimating Naïve Bayesian Classifier Performance

Before analysing the results of the Naïve Bayesian classifier its expected performance was estimated for one evaluation parameter, P_{cc}. The bound on the probability of error, $\mathrm{P}(\epsilon)$, calculated from (3.33) can easily be converted into the probability of correct classification, P_{cc}, as $P_{cc} = 1 - \mathrm{P}(\epsilon)$. Fig. 7.3 shows the variation of P_{cc} with percentage variance and frame duration for the three naïve Bayesian datasets. The dataset correlation is not included since it only affects the size of the reference dataset and therefore the accuracy of the estimates of the mean and covariance. For maximum accuracy the values are calculated from the 0.9 correlation (largest) dataset. The incomplete surfaces in Fig. 7.3 are a result of the determinates in (3.34) tending to infinity preventing calculation of $\mathrm{P}(\epsilon)$. Despite this the general trends are still clear.

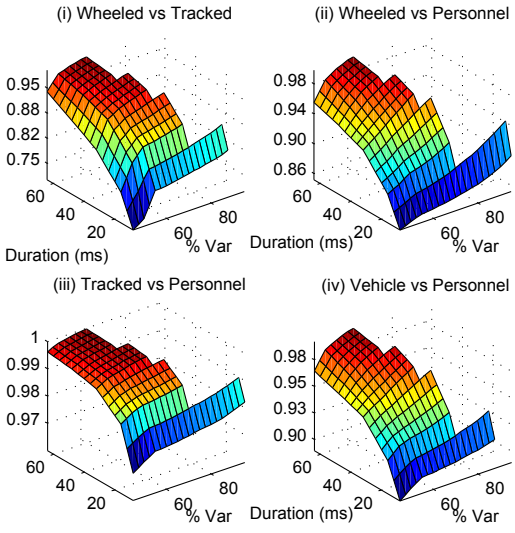
It is apparent that longer frames and more variance give higher probabilities of correct classification. The figures all show that the highest predicted performance was obtained when distinguishing between personnel and tracked vehicles, while the worst predicted performance was when deciding between the two vehicle classes. Increasing the number of number of aspect angles in the dataset appears to reduce the quality of classifier performance. In Fig. 7.3 part (a) the P_{cc} surface has a large plateau of 1.0. The size of this plateau (allowing for the incomplete nature of the surface) is reduced in part (b) and it appears, in part (c), that the plateau would be very small when all the aspects of the targets are included in the reference dataset. Despite this result it is not reasonable to assume that lowering the number of aspects in the reference dataset improves the performance of the naïve Bayesian classifier for two reasons: firstly, the assessment of performance made here does not allow for classifier generalization where extra reference data aspects are expected to improve performance; and secondly, the inclusion of extra aspects in the reference dataset will alter the form of the multivariate probability distributions for the classes and reduce the validity of the Gaussian assumption thus making the Bhattacharyya bound unsuitable.

More detail of the performance of the naïve Bayesian classifier will be given in the following sub-sections.



(a) Bayes single aspect dataset.

(b) Bayes two aspects dataset.



(c) Bayes all aspects dataset.

Figure 7.3: The variation of predicted P_{cc} with percentage variance and frame duration.

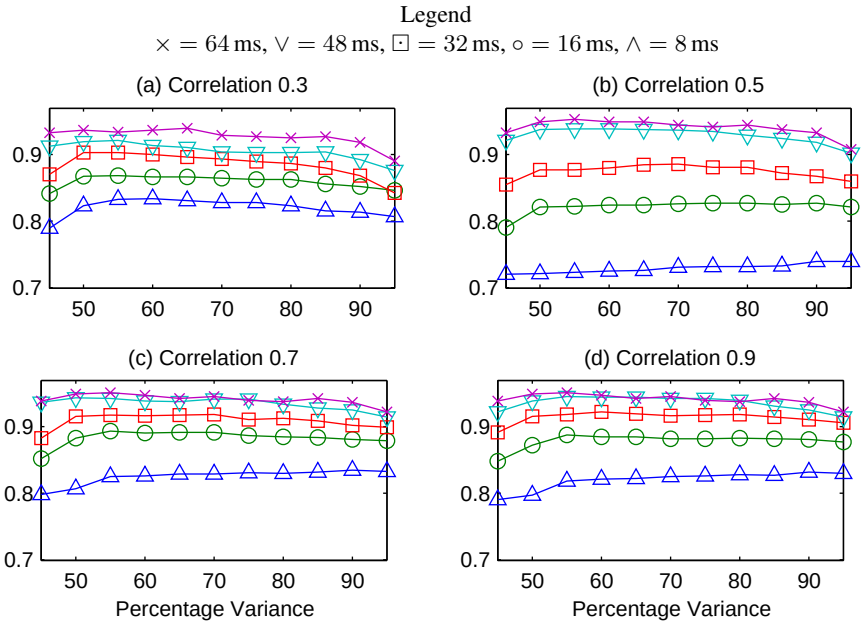


Figure 7.4: Change in P_{cc} with percentage variance.

7.4.3 Naïve Bayesian Classifier Performance

The analysis of the naïve Bayesian classifier is a more complex than the template based classifiers as the feature extraction process can have an impact on the results. During the PCA based dimensionality reduction elements of the feature vectors are discarded and this will inevitably impact the final result. It therefore follows that the first part of the analysis of this classifier is given over to investigating which percentage of total variance should be used. Following that the "unknown" threshold is selected and the performance analysed as was done for the previous classifiers.

Selecting The Percentage Variance

Section 5.4 describes how a percentage of the total data variance can be selected during the PCA process to decide how many elements of the feature vectors should be kept. To evaluate the impact of the percentage variance on the classifier the naïve Bayesian classifier was run against the test data, ignoring the "unknown" and variant targets, and the average probability of correct classification, P_{cc}, calculated.

Fig. 7.4 shows the way in which P_{cc} changes with the percentage variance for the all aspects dataset. Each subplot in the figure shows the situation for a different dataset correlation and each line charts the change in P_{cc} for different frame durations. While it is immediately clear that longer frame durations give higher P_{cc} it is also apparent that a higher percentage variance does not always lead to higher P_{cc}. The value of P_{cc} does increase with the lower percentage variances, from 45% to about 60% depending on case, but after this P_{cc} either remains constant or starts to fall again. Although only the result for the all aspects dataset is shown the trend was observed for the single and two aspects datasets also. It was decided that the percentage variance to be used in classification testing would be that which gave

Dataset	Single Aspect	Two Aspects	All Aspects
Selected Percent Variance	65	55	55

Table 7.5: The percentage variances used in classifier testing.

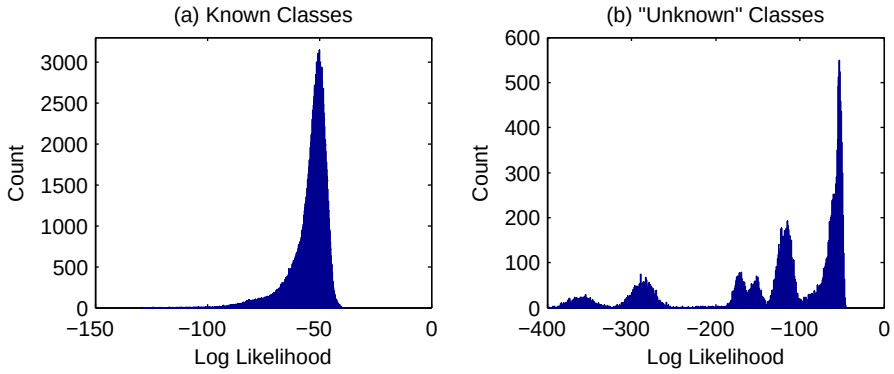


Figure 7.5: Histograms of the naïve Bayesian classifier log likelihoods.

the peak P_{cc} for the 64 ms frame. The choices are listed in Table 7.5 and were made based on the 0.5 correlation dataset.

The reduction in P_{cc} for higher variances was attributed to the increased amount of noise in the high variance feature vectors. As was discussed above, when considering the template classifiers, much of the normalized power spectra used as feature vectors is noise since the μ-DS has a limited bandwidth. The variance in the noise regions of the feature vectors will be small compared to the signature so PCA will give the new basis vectors that representing noise a low score. When low percentage variances are used the noise will be removed from the reduced dimension feature vectors. Since it is considered that noise acts to confuse the classifier it is no surprise that as the percentage variance, and hence noise, increases the performance drops.

With the percentage variance selected the "unknown" threshold could be set.

Setting The "Unknown" Threshold

The purpose of the "unknown" threshold is to detect scores within the classifier that indicate the presented feature vector does not come from a class represented in the reference dataset. In the naïve Bayesian classifier the scores being thresholded are the log likelihoods. A log likelihood of 0.0 represents certainty and less likely events have likelihoods that are increasingly negative. Therefore the log likelihoods of unknown classes should be less than those of the classes the classifier has been trained to recognize.

Fig. 7.5 shows a histogram of the log likelihoods for the single aspect dataset at 65% variance with an 8 ms frame. The dataset correlation isn't included in this analysis since it will have no impact on the result. Changing the correlation changes the ratio of how frames are divided between the test and

Frame Duration (ms)	Single Aspect 65% Min Known	Single Aspect 65% Max Unknown	Two Aspects 55% Min Known	Two Aspects 55% Max Unknown	All Aspects 55% Min Known	All Aspects 55% Max Unknown
8	-120	-21	-151	-40	-129	-46
16	-140	-41	-189	-73	-197	-85
32	-267	-56	-276	-110	-345	-140
48	-301	-68	-277	-132	-405	-190
64	-365	-75	-274	-143	-419	-218

Table 7.6: The limits for the "unknown" threshold level.

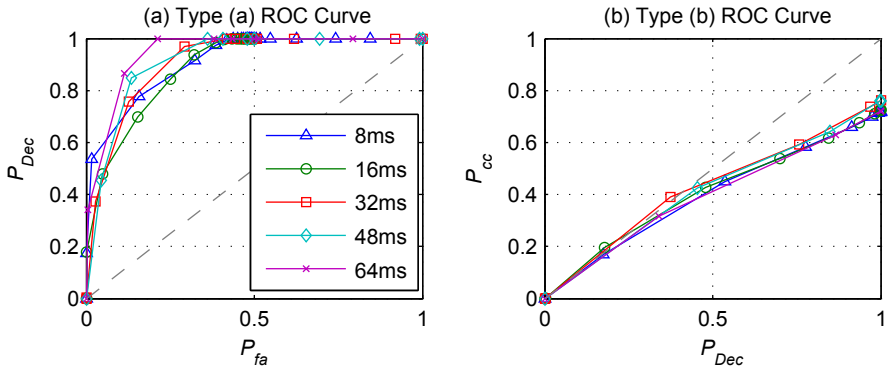


Figure 7.6: The type (a) and type (b) ROC curves for the single aspect naïve Bayesian classifier.

reference datasets; the parameters used to calculate the log likelihoods are estimated from the reference dataset and although changing the number of feature vectors in the reference dataset effects the quality of the estimate it has limited impact on the range of (7.6) compared to the frame duration. Comparing part (a) and (b) of Fig. 7.5 shows that the scores (log likelihoods) for known and unknown classes tend to be different, although there is some overlap. The "unknown" threshold should be varied between the smallest score for a known class and the largest score for an unknown class. Table 7.6 lists the minimum known class and maximum unknown class log likelihoods for the different datasets that define the limits for the "unknown" threshold level. The situation is different to the template based classifiers where the scores, the normalized global distance, were normalized by the duration of the frame and it will no longer be suitable to use one threshold with all frame-duration dataset-correlation parings. It is clear from Table 7.6 that a different threshold will have to be used for different frame durations.

The ROC curves for the three Bayes datasets at the selected percentage variance are shown in Figs. 7.6 to 7.8. In each case thirty possible threshold levels were considered between the limits outlined in Table 7.6; the spacing of the levels was linear. The curves shown are for the dataset correlation level that was selected as giving best performance. This is detailed in Table 7.7 alongside the threshold that was selected for each frame duration. There is a clear difference between these ROC curves and those of the template based classifiers in both the time and frequency domain: the naïve Bayesian approach does not suffer from the step at $P_{\mathrm{FA}} = 0.5$ and this allows a high probability of detection to be achieved at low probability of false alarm. Indeed the false alarm performance of the naïve Bayesian classifier is good

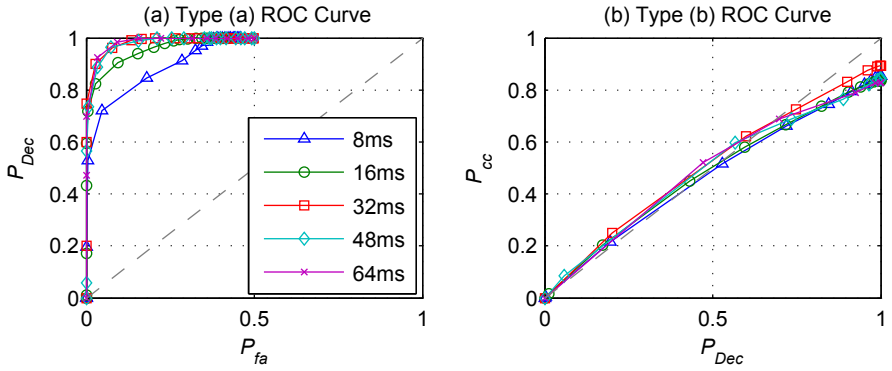


Figure 7.7: The type (a) and type (b) ROC curves for the two aspects naïve Bayesian classifier.

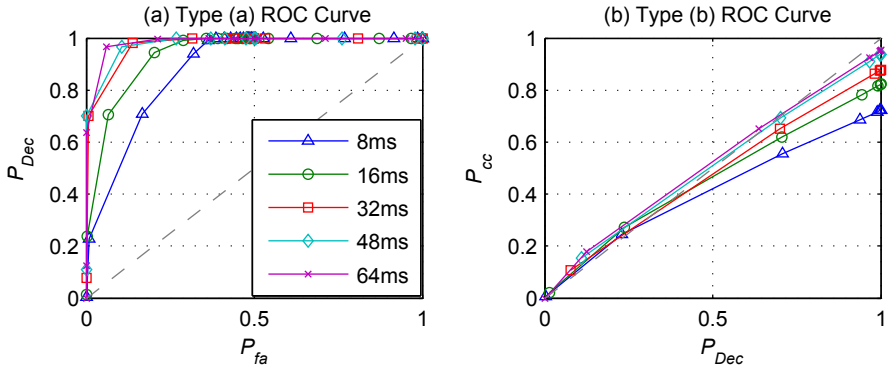


Figure 7.8: The type (a) and type (b) ROC curves for the all aspects naïve Bayesian classifier.

		Threshold for frame duration (ms)				
Dataset	**Correlation**	**8**	**16**	**32**	**48**	**64**
Single Aspect 65%	0.9	-31.2	-52.4	-77.9	-92.7	-105.7
Two Aspects 55%	0.5	-51.8	-72.7	-121.6	-140.8	-165.6
All Aspects 55%	0.9	-30.9	-54.1	-90.0	-118.6	-130.4

Table 7.7: Selected dataset correlation and threshold values for the naïve Bayesian classifier.

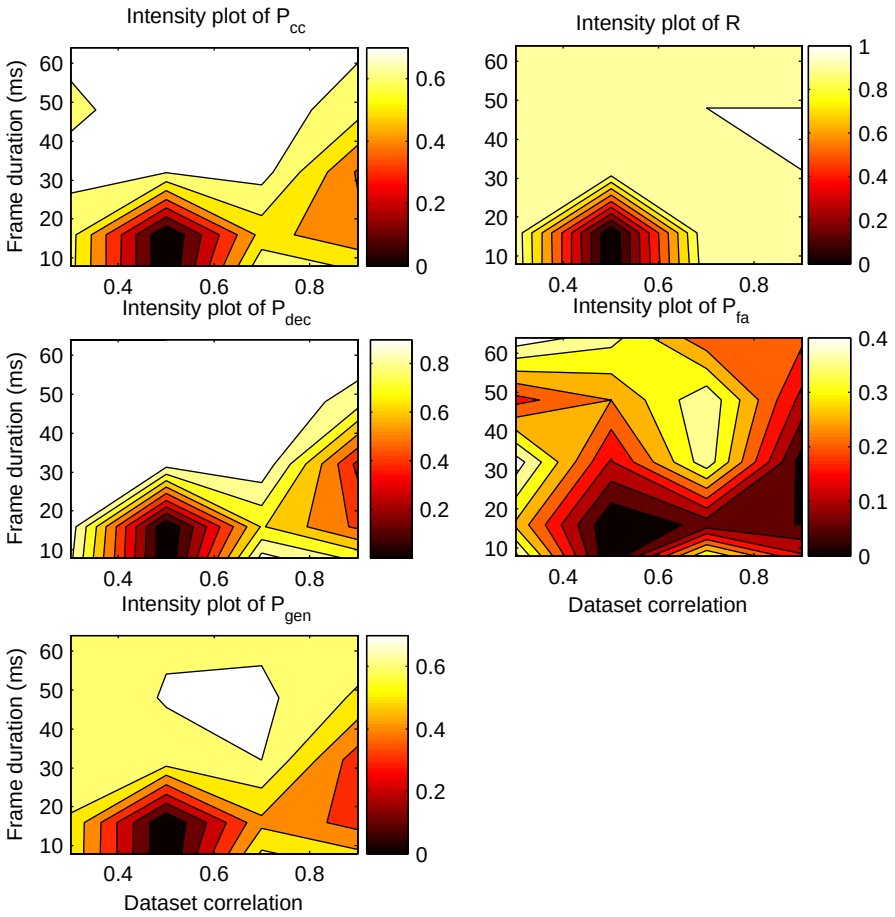


Figure 7.9: Detailed variation of evaluation parameters for the Bayesian classifier with the single aspect 65% variance dataset.

enough that in the two aspect reference dataset case the value of P_{FA}, Fig. 7.7 part (a), never exceeds 0.5 indicating that even for the highest threshold level one of the unknown classes is always successfully detected.

With the level of the "unknown" threshold successfully selected for each dataset and frame duration it was possible to investigate the classifier's performance.

Classifier Performance

The intensity plots showing the variation of the evaluation parameters for the three naïve Bayesian datasets, at the selected percentage variances, are shown in Figs 7.9 to 7.11. The all aspects dataset case does not include an intensity plot for P_{Gen} since the inclusion of all target aspects in the reference dataset means there are no entries in the test dataset with which to assess this parameter. Details of the peak performance of the classifier are provided in Table 7.8: the dataset correlations from Table 7.6 were in use and the frame durations that gave the peak performance is indicated. When comparing these results with those for the template based classifier it is important to remember that here different "unknown" threshold levels are being used for each frame duration so slightly improved P_{Dec} and P_{FA} performance

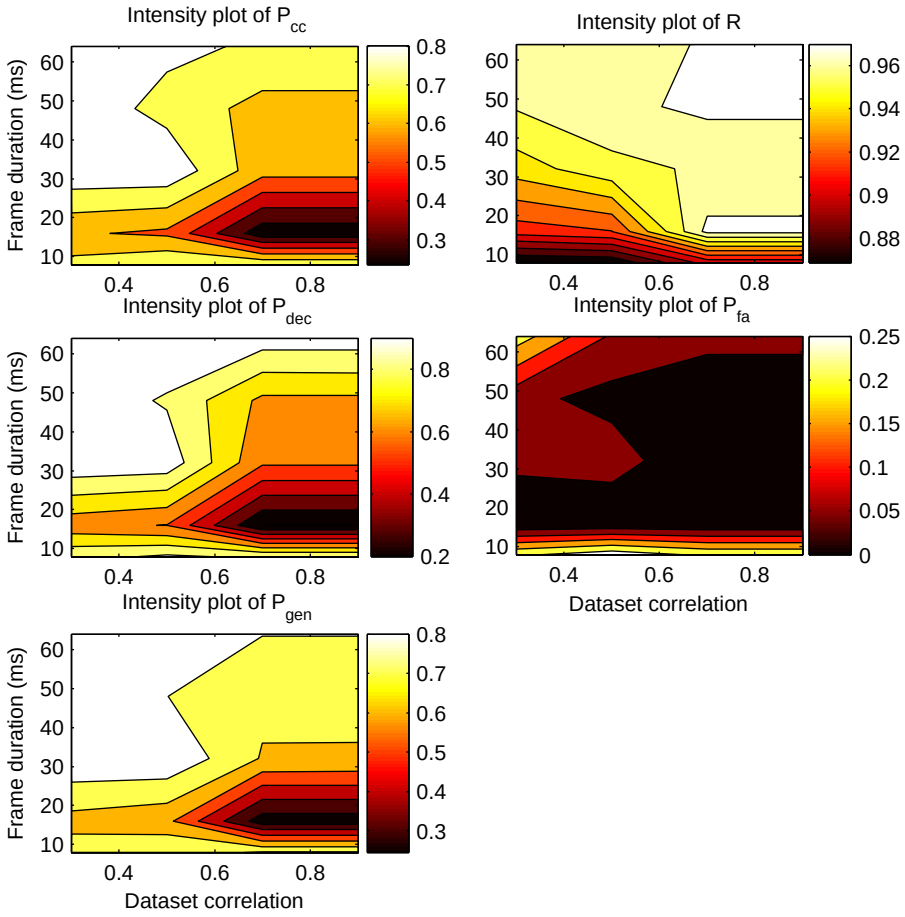


Figure 7.10: Detailed variation of evaluation parameters for the Bayesian classifier with the two aspects 55% variance dataset.

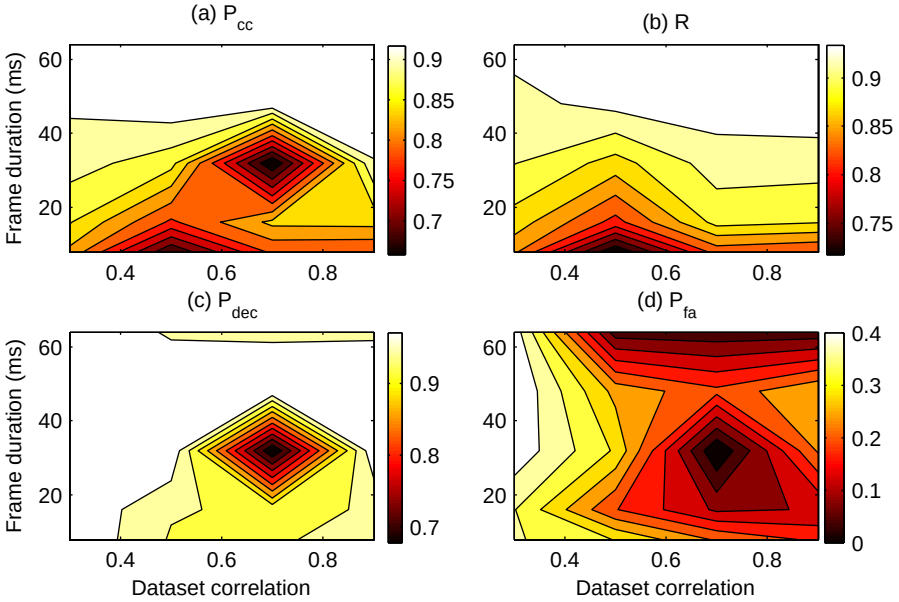


Figure 7.11: Detailed variation of evaluation parameters for the Bayesian classifier with the all aspects 55% variance dataset.

	Single Aspect 65%	**Two Aspects 55%**	**All Aspects 55%**
Frame duration (ms)	64	32	64
P_{cc}	0.72	0.88	0.92
RL	0.99	0.96	0.95
P_{Gen}	0.68	0.88	N/A
P_{FA}	0.21	0.08	0.05
P_{Dec}	0.68	0.96	0.97

Table 7.8: The classifier evaluation parameters for the three naïve Bayesian datasets.

would be expected.

Figs 7.9 to 7.11 show that, in general, the P_{cc} and reliability (RL) for the naïve Bayesian classifier was greatest when the frame duration was longest and dataset correlation highest. That longer frame durations give improved performance fits with earlier observations that longer frames contain more information about the target. The improvement with higher dataset correlation was attributed to the same underlying effect as for the template based classifiers—that more reference examples give better performance—but the practical mechanism is slightly different. For the template classifiers, the higher the dataset correlation the more examples there are for the input to be contrasted with. Alternatively, for the naïve Bayesian classifier the larger the reference dataset the better the estimates of μ_{ij} and σ_{ij}^2 from (7.5) will be leading to better estimates of $\mathrm{L}(x_i|c_j)$ and hence better performance.

The results for P_{FA} and P_{Dec} were most encouraging, especially when compared with the results for the template based classifiers in both the frequency and time domain. The thresholding of the log likelihood scores within the naïve Bayesian classifier worked well and it was possible to simultaneously achieve a good P_{FA} and P_{Dec}. The performance for these parameters increased as the number of aspects in the reference dataset increased and it can be seen in Table 7.8 that the two aspects dataset results are almost the same as those for the all aspects dataset that contain most information about the known classes. Most significantly it is seen that the probability of mistakenly declaring an unknown class as a known falls to less than 0.1 and that the probability of declaring a known class rises above 0.95. Furthermore it is also seen that the probability of generalization, P_{Gen}, is significantly better than an arbitrary declaration.

For a classifier making an arbitrary decision the expected value of P_{Gen} would be 0.33 since any of the three known classes could be randomly selected; the worst case, the single aspects dataset, with the naïve Bayesian classifier gives $P_{\text{Gen}} = 0.68$. The inclusion of a second aspect in the reference dataset increases this value to 0.88. It was not possible to measure P_{Gen} for the all aspects case since the inclusion of all target aspects in the reference dataset left none to use as variant targets. This performance is a significant improvement over the template based methods where the highest P_{Gen} obtained was 0.66 in the frequency domain and 0.53 in the time domain. This improvement is attributed to the removal of noise elements that increase the similarity of the data frames from the feature vector by PCA.

Input Class	Declared as... Wheeled	Tracked	Personnel	Unknown
Wheeled Vehicle	0.97	0.07	0.03	0.01
Tracked Vehicle	0.01	0.83	0.00	0.05
Personnel	0.02	0.01	0.97	0.00
"Unknown"	0.00	0.09	0.00	0.95

Table 7.9: Reduced confusion matrix for Bayes all aspects 55% dataset.

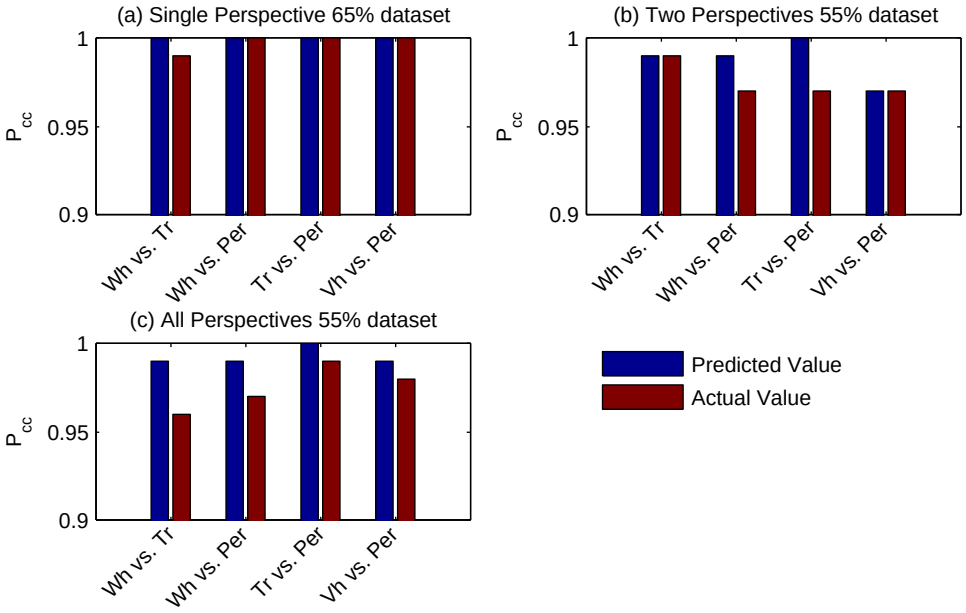


Figure 7.12: Comparison of the predicted and actual performance of the naïve Bayesian classifier.

The last part of the naïve Bayesian classifier analysis is to contrast its actual performance with that predicted by the Bhattacharyya error bound. To make this comparison the pairwise classifications for which P_{cc} was estimated must be derived from the classifier results, this was achieved via calculation of the pairwise $\mathrm{P}(\epsilon)$. Table 7.9 shows the confusion matrix for the all aspects 55% dataset. In the table the different input angles have been merged together e.g. the input marked as "Wheeled Vehicle" is an average of all aspect angles that were available for the wheeled vehicle class in the test dataset. To calculate the $\mathrm{P}(\epsilon)$ for a particular pairing of targets further averaging is used. For example, the $\mathrm{P}(\epsilon)$ in the binary decision between the wheeled and tracked vehicles is the mean of the instances in the confusion matrix when wheeled vehicles were mistaken for tracked vehicles and when tracked vehicles were mistaken for wheeled vehicles. In this instance that is $\mathrm{P}(\epsilon) = \frac{0.07+0.01}{2} = 0.04$. The P_{cc} for the binary decision is then $1 - \mathrm{P}(\epsilon) = 1 - 0.04 = 0.96$. By repeating the process for the other pairings considered during performance prediction it is possible to calculate the actual pairwise errors to compare with those generated by the Bhattacharyya bound. The actual P_{cc}'s for the binary decision are detailed alongside their predictions for each dataset in Fig. 7.12. The frame duration and dataset correlations used are those listed above as giving best performance.

The results of Fig. 7.12 suggest that despite the assumptions made during the performance prediction the estimates and actual results are in reasonable agreement. The *y*-scale of the bar charts is very small, showing the interval [0.9, 1.0], so despite some discrepancies between actual and predicted performance it is never significant. When the number of aspects in the reference dataset is low the predictions are most accurate with only a slight deviation for the wheeled vs. tracked vehicle decision. In the all aspects dataset the actual performance lags the prediction slightly in all the pairings although the trends in the results are the same. This increase in discrepancy between prediction and actual performance with number of aspects was attributed to the increase in complexity of the problem. The Bhattacharyya estimate of $\mathrm{P}(\epsilon)$ is based solely upon the reference dataset where no examples of the variant entries from the test dataset are available. Therefore, when the actual error probabilities were calculated the inputs for variant target were ignored and only those classes in the reference data were included in the calculations. As a result the complexity of the binary decisions increase with the number of aspects in the reference dataset since more aspects cause to more variation in the feature vectors. That there is more discrepancy between predictions and reality for more complex problems is not surprising.

Furthermore, it is to be expected that the predictions of performance give more optimistic results. The method for predicted $\mathrm{P}(\epsilon)$, and hence P_{cc}, from section 3.6 are for binary decisions but the naïve Bayesian classifier is not making a binary decision. Perhaps more significant is that the prediction takes no account of the "unknown" detection threshold. In Table 7.9 when wheeled inputs are presented "unknown" is declared with a rate of 0.01 while when tracked inputs are presented the "unknown" rate is 0.05. This means that before there is any confusion between the two classes in a binary decision the rate of predicted P_{cc} *must* be lower as 3% of the inputs have been rejected as "unknown" making a P_{cc} of 1.00 impossible. Despite these limitations the use of the Bhattacharyya bound on $\mathrm{P}(\epsilon)$ has been demonstrated as providing a useful prediction of the performance of the naïve Bayesian classifier.

With the analysis of the naïve Bayesian classifier completed it is possible to contrast the performance of the different frequency domain classifiers investigated.

7.4.4 Comparison Of Classifiers And Discussion Of Limitations

In contrast to what had been expected, the template based classifiers did not fair particularly well operating in the frequency domain. Their performance was either the same as, or slightly worse than, the time domain versions depending on the number of aspects stored in the reference dataset. Conversely, the results for the naïve Bayesian classifier, that include the complete feature extraction process, demonstrate that high classifier performance can be obtained in the frequency domain using this method.

The key indicator that there was a problem with the template based approaches was that performance improved as frame duration fell. The move to the frequency domain had been expected to make the situation easier for the template classifiers, by removing the problem of μ-DS starting phase that had

so troubled time domain k-NN. Unfortunately, the decision to use the template classifiers as an intermediate step in the development of the full feature extraction process meant that while the problem of the starting phase was removed other difficulties were introduced. However, the advantage of operating in the frequency domain was demonstrated by the naïve Bayesian classifier.

With the full feature extraction process in place and the use of a naïve Bayesian classifier impressive performances were obtained. As had been demonstrated by the template classifiers the use of a dB scale for the feature vectors assisted classification while the application of PCA for dimension reduction removed the noise parts of the vectors that had confused the template classifiers. With this extra pre-processing the naïve Bayesian approach was able to demonstrate that it is possible to use the μ-DS to classify radar targets accurately and reliably. Furthermore, with the pre-processing in place the "unknown" rejection threshold was able to successfully detect the majority of the unknown target class inputs without significantly effecting the classifier's rate of declaration. This was indicated by low value of P_{FA} and the high value of P_{Dec}. In addition to good basic performance the naïve Bayesian classifier also excelled in generalization.

With the exception of the time domain k-NN, all of the classifiers considered have made better than arbitrary decisions for the variant input classes. For the frequency domain template classifiers P_{Gen}'s of ≈ 0.55 for the single aspect and ≈ 0.65 for the two aspects reference dataset were obtained. Despite this being better than an arbitrary decision it is still much lower than the target level of 0.8 that has been identified as indicating acceptable performance for radar Automatic Target Recognition (ATR) (Stove, 2006). For the naïve Bayesian classifier using the single aspect dataset the value of P_{Gen} obtained was 0.68 and the inclusion of a second aspect in the reference data caused this value to rise to 0.88, which is above the acceptable performance level. It was not possible to assess P_{Gen} when the all aspects dataset was used, since it leaves no aspects to be regarded as variants, but if the relation of P_{Gen} to number of aspects is the same as for the other evaluation parameters then the two aspect performance will be close to that for all aspects. Despite these success there was still a significant limitation on the naïve Bayesian classifier analysis.

The selection process for the "unknown" threshold did not allow for each frame duration having its own set of possible values when the naïve Bayesian classifier was being tested. For each frame duration a set of thirty possible levels, each with and index, for the threshold were generated, the classifier was run with each threshold and the ROC curves were generated. The index of the threshold that gave best performance over all frame durations and dataset correlation levels was then selected by analysing the ROC curves. The same index was then used to select the threshold level at all frame durations: it was not realized until after the testing was completed that there was no guarantee that the same index gave best performance for each frame duration. Fortunately this oversight means that the reported results, for frame durations other than that which gave best performance, are pessimistic. Had the best threshold

level for each duration been chosen then the value of the evaluation parameters in Figs. 7.9 to 7.11 would have been improved. It is important to stress that this limit has no impact on the results of Table 7.8 since this details the results for the frame duration with which the threshold index was selected.

The principal limitation for the template based classifiers was the feature vector pre-processing. As was discussed above they were unable to operate effectively with the high variance in the amplitudes of the feature vectors and the large noise regions. A valuable extension to the work presented here, therefore, would be to repeat the template based investigation but using the full feature vector extraction process.

The work of this chapter has demonstrated the μ-DS is a suitable phenomenon for use in distinguishing radar targets. The results of the naïve Bayesian classifier are comparable, in performance terms, with those of (Stove and Sykes, 2003; Stove, 2006) that are for a commercially available radar ATR solution. Following this work the focus of the investigation shifted to investigate how changing the nature of the radar from monostatic to bi- or multistatic affected the μ-DS.

7.5 Summary

By working in the frequency domain, rather than the time domain, it was possible to implement the full feature extraction process described in Chapter 5. Furthermore, it became possible to implement a third classifier: the naïve Bayesian. The two template based approaches, k-NN and DTW only required modification to their feature extraction routines to work in the new domain. To make the changes between the frequency domain and the time domain work more gradual, the template based classifier were used with a limited version of the feature extraction process in which no feature vectors dimension reduction was performed i.e. the PCA processing was omitted.

The results for the template based classifiers were disappointing. It had been anticipated that it would be easier to perform radar ATR in the frequency domain, but the k-NN and DTW results did not support this expectation. The classifiers were better able to generalize in the frequency domain, but for single aspect reference datasets the overall performance was slightly worse than the time domain while for two aspect datasets the results were the same as the time domain. Furthermore, in contrast to the general literature is was observed that the performance of both k-NN and DTW dropped as frame duration increased. Detailed study of the workings of the classifiers showed this to be a result of the transition to the frequency domain limiting the μ-DS to be in a few elements of the feature vector rather than spread throughout it as had been the case in the time domain. In contrast to these disappoint results, however, the naïve Bayesian classifier had a particularly impressive performance.

The performance of the naïve Bayesian classifier, which operated with the full feature extraction process, varied depending on the frame duration, dataset correlation, number of target aspects in the reference dataset and the amount of variance remaining in the feature vectors following their extraction.

In contrast to the performance of the template based classifiers, longer frames resulted in better performance for the naïve Bayesian classifier. Again, as expected, high levels of dataset correlation gave best performance. However, the probability of correct classification was seen to peak when between 45% and 60% of variance remained in the feature vector. As higher percentages were used the level of P_{cc} either remained constant or began to fall. The result was attributed to increasing amounts of noise passing the PCA process when higher percentage variance was used. When the number of target aspects included in the reference data was increased from one, dramatic performance improvements were seen. By moving to a two aspect dataset, the value of P_{cc} rose from 0.72 to 0.88 and the generalization, P_{Gen}, rose from 0.68 to 0.88. Tests for inclusion of all available target aspects in the reference dataset were also conducted and it was seen that P_{cc} rose to 0.92 suggesting the majority of the gain was in the inclusion of the second aspect. The results for the other performance metric were also impressive and indicated the classifier was capable of detecting "unknown" inputs correctly. The results obtained for the naïve Bayesian classifier's P_{cc} matched well with those predicted using the Bhattacharyya bound on the error.

By averaging entries of the confusion matrices produced during classifier testing it was possible to derive a P_{cc} between pairs of targets that were contrasted with the performance predictions. The comparison showed that the classifier was operating at close to the performance predicted by the method from Chapter 3. This result was regarded as particularly significant since a number of assumption had been made in formulating the prediction method and it was thought these might limit its success. However, the method was actually demonstrated as giving good estimates of P_{cc} and being able to show trends in the classifier performance. This successful result marked a novel extension to the methodology of radar ATR evaluation.

With the frequency domain testing completed, the next subject to be investigated was the use of multistatic radar data to form a multiperspective radar ATR system.

Chapter 8

Multistatic Micro-Doppler

Multiperspective classification methods have been shown to improve the performance of radar Automatic Target Recognition (ATR) systems by simultaneously allowing the classifier access to information about the target from different views. A method for obtaining simultaneous multiperspective views is to use a multistatic rather than a monostatic radar. To use such an approach for Micro-Doppler Signature (μ-DS) based radar ATR required the development of a multistatic micro-Doppler theory: such a theory is developed in this chapter.

With the theory in place a set of multistatic field trials are described and the results from those trials analysed. Field trial data is then used in the first reported attempt at multiperspective μ-DS based radar ATR. While the results are unexpectedly poor, it is apparent that this is a result of the nature of the radar used to gather the data rather than the concepts itself. The chapter is closed with a summary that identifying where the future advances to multiperspective μ-DS classification are expected.

8.1 Monostatic Micro-Doppler Signature Limitations

The Micro-Doppler Signature (μ-DS) of a radar target has a strong dependency on the target aspect angle and this can limit the performance of a radar Automatic Target Recognition (ATR) system based on the signature. Section 4.4 showed series of figures, Figs. 4.8 to 4.13, that contain the spectrograms of the μ-DS for the three target classes used in the testing of chapters 5 and 7 at different aspect angles. In section 4.4 it is discussed how changing the aspect angle affects the μ-DS with the following aspect angle dependencies identified:

- At some aspect angles certain parts of the μ-DS are not visible due to occlusion of the moving parts of the target by other elements of the target. A phenomena known as *self occlusion*.
- The absolute frequency of the Main Doppler Line (MDL) of the signature is at a maximum for the head-on (0°) and tail-on (180°) to the radar Line Of Sight (LOS) aspect angles. The absolute

frequency is lower for all other aspects.

- The bandwidth of the MDL of the signature is at a maximum for the head-on (0°) and tail-on (180°) to the radar LOS aspect angles. The bandwidth is lower for all other aspects.
- The distance between the spectral lines is dependent on bandwidth. At low bandwidths the lines merge together reducing the ability to resolve features in the signature.

So far the affect of μ-DS aspect dependency has been investigated by changing the number of aspects of the target included in the reference dataset. However, alternative approaches may be taken based on combining multiple measurements of the target taken at different aspects—as multiperspective approach.

The difficulty of aspect dependency of a target signature has been encountered in other radar ATR situations, such as High Range Resolution Profiles (HRRPs) (Liao et al., 2000), and has been overcome by considering multiple target aspects during each classification (Zhu et al., 2007; Vespe, 2007). In the former of these studies no indication of how the multiple target aspects are obtained is given, although the figures of the paper suggest it is as a result of target motion relative to the radar. In the latter study the HRRPs are taken from an Inverse Synthetic Aperture Radar (ISAR) image of the target so multiple aspects are certainly obtained as a result of target motion. The two studies use very different methods to combine the perspectives with Zhu using a Hidden Markov Model (HMM) while Vespe uses various approaches, from concatenation to multiplication of likelihoods for individual perspectives, depending on the classifier being used. The Dynamic Time Warping (DTW) classification technique that has already been investigated in this study is capable of processing data series where each sample is multidimensional suggesting that it would able to process a multiperspective input vector. Although before considering such a classifier it will be necessary to define the multiperspective μ-DS.

The approach taken in (Zhu et al., 2007; Vespe, 2007) of using the target motion to obtain separation between different aspects of the target will not be suitable for the μ-DS. Inherently, if a single sensor is employed and target motion used to obtain aspect separation then the outputs will be separated in time. Since the μ-DS is a time dependent phenomenon, see section 4.1, the different target aspects need to be gathered simultaneously or they will be mismatched and unsuitable for data fusion. Such a dataset might be gathered using a multistatic radar system.

8.2 Monostatic, Bistatic And Multistatic Radar

The most common form of radar is the monostatic radar, where both the transmitter and receiver are co-located, but there are other possible configurations. In general a monostatic radar has a single antenna that is shared by both the transmitter and the receiver, but the two channels may have separate antennas as shown in Fig. 2.1. (Such separation was included by design in some early Continuous Wave (CW) radar to reduce the feed through between the two channels.) When the distance between the transmitter and receiver antenna is small compared to the target range the antennas may be considered co-located

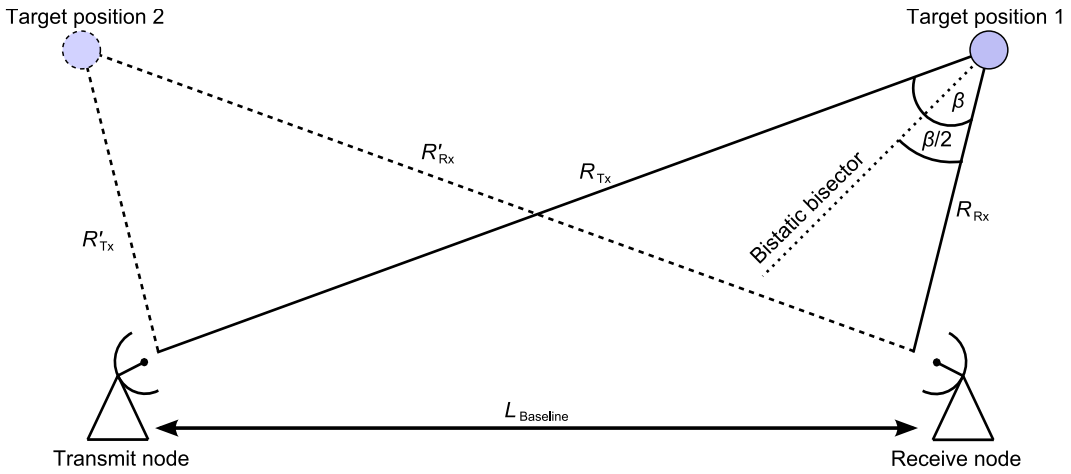


Figure 8.1: The topology of a bistatic radar.

and effectively the same as a shared antenna, although the properties of the two antennas may differ. If the distance between the antennas is significant when compared to the target range then the radar is a *bistatic radar* meaning that the transmitter and receiver are considered as being at different locations. It is also possible for a radar to be *multistatic* where there are multiple transmitters and receivers spread over many different locations (Chernyak, 1998). Under a multistatic topology an individual site, or node, in the network may be capable of both transmitting and receiving. Such networks can be considered to as a collection of bistatic channels so here focus is given to the difference between monostatic and bistatic systems.

The topology of the bistatic radar is shown in Fig. 8.1. The line between the two nodes is referred to as the *bistatic baseline* and it has a length L_{Baseline}. The lines from the transmitter and receiver to the target are the *transmitter line segment* and *receiver line segment* respectively. Together the three lines make up the *bistatic triangle* (Cherniakov, 2007). The angle between the transmitter and receiver line segment is the *bistatic angle*, β, and this angle is bisected by the *bistatic bisector*. The *bistatic range* to target position 1 in the diagram, R_{bistatic}, is the sum of the range to the target from the two nodes, i.e.

$$R_{\mathrm{bistatic}} = R_{\mathrm{Tx}} + R_{\mathrm{Rx}}, \tag{8.1}$$

and this range can be found from the time delay, Δt, between signal transmission and reception according to:

$$R_{\mathrm{bistatic}} = c\Delta t. \tag{8.2}$$

In the monostatic case $R_{\mathrm{Tx}} = R_{\mathrm{Rx}}$ so it was possible to halve the $c\Delta t$ term and find the true target range (see (2.1)) but when $R_{\mathrm{Tx}} \neq R_{\mathrm{Rx}}$ this is not possible and there is inherent ambiguity in the target's location. In Fig. 8.1 a second set of ranges, R'_{Tx} and R'_{Rx} are shown with dashed lines. The sum of these ranges is the same as sum of R_{Tx} and R_{Rx} so it would not be possible to determine if the target

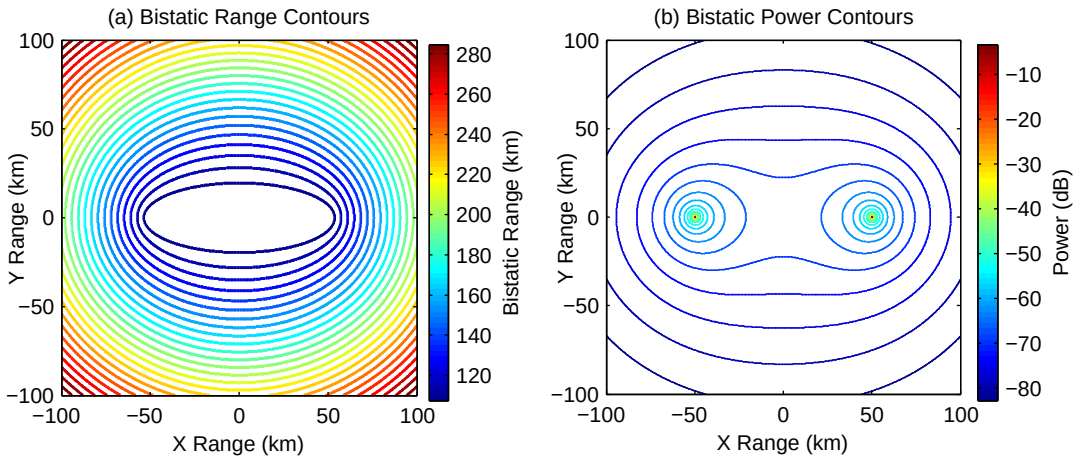


Figure 8.2: The contours associated with bistatic radar configurations for (a) range and (b) power.

were at position 1 or 2 through the use of (8.2) alone. For any given R_{bistatic} the location of the target is constrained only to an ellipse with foci at the position of the two nodes. This situation is shown in Fig. 8.2 part (a). The nodes are positioned at the points (-50, 0) and (50, 0) and the contour lines indicate where the bistatic range is equal. If both nodes were at the same position the radar would be monostatic and the contour lines would be circular. The situation is similar for the received power in a bistatic radar.

The power received is given by the bistatic radar equation as (Willis, 2005):

$$P_{\text{rx}} = \frac{P_{\text{tx}} G_{\text{tx}} G_{\text{rx}} \sigma_{\text{bi}} \lambda^2 F_{\text{tx}}^2 F_{\text{rx}}^2}{(4\pi)^3 (R_{\text{tx}} R_{\text{rx}})^2 L_{\text{tx}} L_{\text{rx}}} \tag{8.3}$$

where P_{rx} and P_{tx} are the transmitted and received power, G_{tx} and G_{rx} are the gains of the transmit and receive antenna, σ_{bi} is the *bistatic* Radar Cross Section (RCS) of the target, F_{tx}^2 and F_{rx}^2 the pattern propagation factors for the transmit and receive paths, λ the carrier wavelength, R_{tx} and R_{rx} the range from the transmit and receive antennas to the target, and L_{tx} and L_{rx} the losses in the transmit and receive channel. The principal differences between the bistatic and the monostatic radar equation (see (2.3) for the monostatic version) are the splitting of the range, pattern propagation factor and system losses into two values, one for the transmit and one for the receive node, and the use of the bistatic RCS. Fig. 8.2 part (b) is a contour plot showing the lines of constant power when the radar nodes are at positions (-50,0) and (50,0). Again, if the receiver and transmitter are colocated the contour pattern is circular: the result for a monostatic radar. In addition to having different power and range equations bistatic radars also have a specific Doppler equation.

The Doppler shift in a bistatic configuration depends on the angle, ϕ, between the velocity vector, $\mathbf{v}$, and the bistatic bisector as well as the velocity and carrier wavelength. Fig. 8.3 shows the situation when a moving target is illuminated by a bistatic radar. The Doppler shift may be calculated as (Cherniakov,

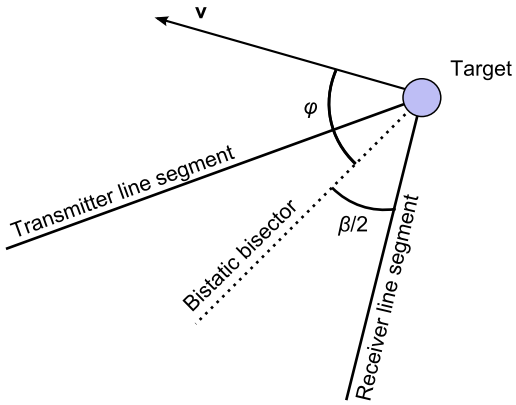


Figure 8.3: The configuration for a bistatic radar illuminating a moving target.

2007):

$$f_{\mathrm{bi-Dop}} = -\frac{2|\mathbf{v}|}{\lambda}\cos(\phi)\cos\left(\frac{\beta}{2}\right). \tag{8.4}$$

The $\cos(\phi)$ term effectively projects the velocity along the bistatic bisector so the bistatic Doppler can be considered a function of the target's bisector velocity in the same way that the monostatic Doppler is a function of the target's radial velocity. Equation (8.4) indicates that bistatic Doppler shifts will always be smaller than their monostatic counterparts. If the transmit and receive nodes are co-located then $\beta = 0$ and $\cos(\frac{\beta}{2}) = 1$ but for all other values of β the $\cos(\cdot)$ term will be less than one reducing the value of $f_{\mathrm{bi-Dop}}$ relative to the monostatic case. This is of particular significance to the study of μ-DS since the bandwidth of the signature depends on the maximum Doppler shift that may be induced. It is therefore clear that the μ-DS bandwidth will always be lower in bistatic systems when compared to monostatic ones.

At the start of this section the concept of multistatic radar was introduced and it was stated that the multistatic radar could be considered as a number of bistatic channels. In Fig. 8.4 a schematic for a small multistatic network is shown in which there are two transmit and receive nodes, Nodes 1 and 3, and a single receiving node, Node 2. The different possible bistatic baselines are indicated with solid lines and the various paths taken by the radar transmissions are indicated with dotted lines. There are six possible channels that may be formed in the network: two monostatic when Node 1 and Node 3 receive their own transmissions; two bistatic on baseline $\mathrm{L}^1_{\mathrm{Baseline}}$ when Node 1 and Node 3 receive each others transmissions; one on baseline $\mathrm{L}^2_{\mathrm{Baseline}}$ when Node 2 receives Node 1's transmission; and one on baseline $\mathrm{L}^3_{\mathrm{Baseline}}$ when Node 2 receives Node 3's transmission. In general the number of channels in the multistatic network is:

$$\text{Num. Channels} = \text{Num. Transmitters} \times \text{Num. Receivers} \tag{8.5}$$

By suitably encoding the transmissions it is possible to operate all of the channels simultaneously (Deng,

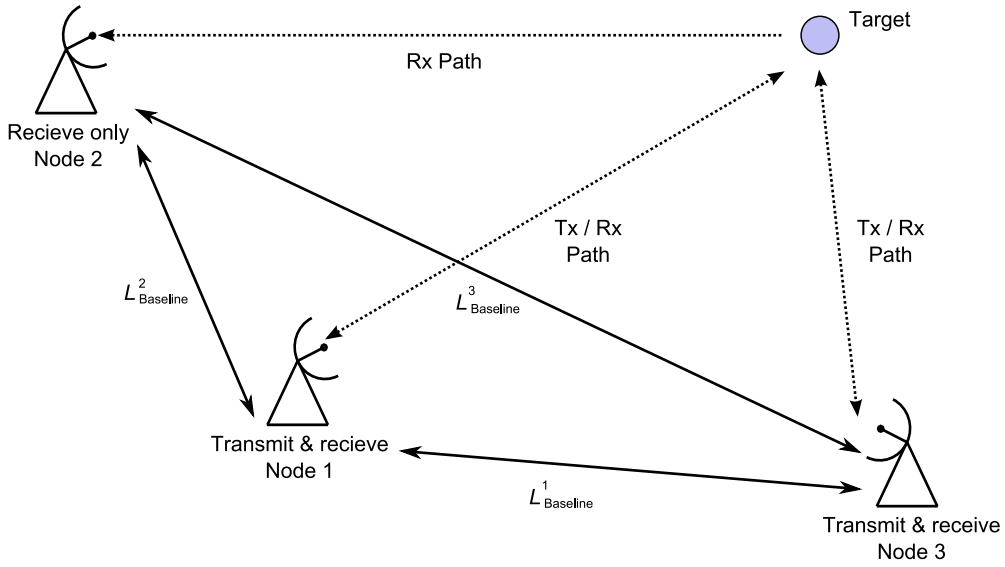


Figure 8.4: A multistatic radar network.

2004). Since each channel will provide a different aspect of the target such a network could provide the simultaneous multistatic μ-DS that was proposed at the end of section 8.1.

Before moving on to investigate real examples of a multistatic μ-DS, that were gathered at a field trial using University College London (UCL)'s NetRAD system, it is necessary to develop a theory for the multistatic μ-DS. Using the theory a model for one of the targets used during field trials, personnel, was developed allowing comparison of theoretical and real results.

8.3 The Multistatic μ-DS

The μ-DS for simple types of motion has already been modelled for the monostatic case (Chen et al., 2006), see section 4.1. The process is to consider the target as a series of sub targets, work out the signal that is received from each of these and then coherently sum the signals. Calculation of the signals is achieved by modelling the change in position of the sub targets with time and then deriving the phase change of the appropriate signal. The baseband signal is then:

$$u(t) = \sum_{p=1}^{N} \sqrt{P_p^{\text{rx}}} \frac{\mathrm{d}}{\mathrm{d}t} e^{-i\phi_p(t)} \tag{8.6}$$

where P_p^{rx} is the power of the signal for the p^{th} subtarget calculated from the radar range equation (Skolnik, 1980) and $\phi_p(t)$ is the phase expression for the p^{th} subtarget. The phase expression is:

$$\phi_p(t) = \begin{cases} \frac{2\pi.2R_p^{\text{mono}}(t)}{\lambda_{\text{c}}} & \text{for monostatic channels} \\ \frac{2\pi.R_p^{\text{bi}}(t)}{\lambda_{\text{c}}} & \text{for bistatic channels} \end{cases} \tag{8.7}$$

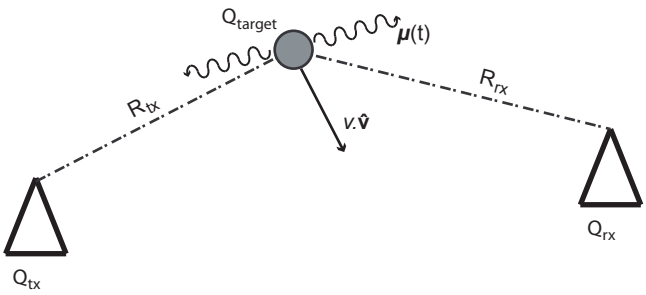


Figure 8.5: The radar-target geometry of a bistatic channel.

where $R_p^{\text{mono}}(t)$ is the monostatic range and $R_p^{\text{bi}}(t)$ is the bistatic range to the p^{th} subtarget; λ_{c} is the carrier signal wavelength. (Note that the factor of 2 is *not* required for the bistatic case since the two way range, rather than one way, range is used.) In the following attention is given to a single bistatic channel since the multistatic case is a collection of bistatic channels, see section 8.2.

Fig. 8.5 shows a general radar-target geometry, with a target exhibiting a micro-motion, for a bistatic channel. The transmit node is at point Q_{tx}; the receive node, Q_{rx}; and the target, $Q_{\text{target}}(t)$. The target has speed v in a direction $\hat{\mathbf{v}}$. Additionally the target has a micro-motion that is described by the expression $\boldsymbol{\mu}(t)$ s. The position of the target can then be written:

$$Q_{\text{target}}(t) = Q_{\text{target}_0} + \hat{\mathbf{v}}vt + \boldsymbol{\mu}(t) \tag{8.8}$$

where Q_{target_0} is the position of the target at $t = 0$s. The bistatic range is then:

$$\begin{aligned} R_{\text{bistatic}}(t) &= R_{\text{tx}}(t) + R_{\text{rx}}(t) \\ &= \|Q_{\text{target}}(t) - Q_{\text{tx}}\| + \|Q_{\text{target}}(t) - Q_{\text{rx}}\| \end{aligned} \tag{8.9}$$

where $R_{\text{tx}}(t)$ and $R_{\text{rx}}(t)$ are the ranges from the transmit and receive node and $\|\cdot\|$ is the Euclidean norm. Substituting $R_{\text{bistatic}}(t)$ for each subtarget into (8.7) will provide expressions for phase that can be used to calculate the baseband signal, including the μ-DS, using (8.6). The multistatic case is a collection of $N_{\text{tx}} \times N_{\text{rx}}$ bistatic channels, formed by pairing transmit and receive nodes, each described by (8.6). N_{tx} and N_{rx} are the number of transmit and receive nodes respectively.

It is the number of channels available that differentiates the multistatic μ-DS from the monostatic μ-DS. In the monostatic case there is a single channel that relates to a single target aspect while in the multistatic case there are several channels relating to multiple aspects of the target. However, this gain in information of the target does not come for free. In the monostatic case the signature depends solely on the target aspect. In the multistatic case the signature depends not only on target aspect but also the radar network topology that will determine two significant features of the multistatic μ-DS: the number of channels that make up the signature and the angular separation between these channels. As will be

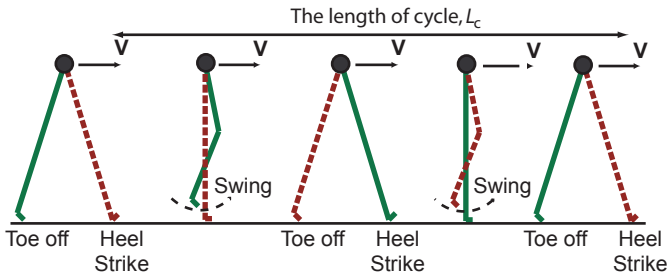


Figure 8.6: The cycle of human locomotion.

seen in section 8.4 there is a correlation between the channels; this means the amount of information available on the target does not increase linearly with the number of channels: i.e. doubling the number of channels does not necessarily double the amount of target information obtained. To investigate this it is necessary to model a specific target.

To model a specific target two further parameters are required: the number of scatterers to include in the model; and the nature of the micro-motion, $\boldsymbol{\mu}(t)$, associated with each scatterer. In the next section a simplified model of human locomotion is developed using three scatterers. The micro-motions are based on the formulae used in computer animations of moving people.

8.4 Modelling The Multistatic μ-DS Of Personnel

The movement of a person's body during locomotion has been modelled in great detail to assist in computer animation (Boulic, 1990). Such models have been used to make estimates of the monostatic μ-DS and as the basis for ATR feature extraction (van Dorp and Groen, 2003). Here a greatly simplified model of human locomotion is presented; despite its simplicity it is still consistent with the detailed model of (Boulic, 1990).

A personnel target was considered as comprising of three parts: a torso and two legs. Arms were not considered since they swing in phase with the legs and in the simple model developed they were considered to merely increase the RCS of the legs. The relative motion of these three parts during a single cycle of walking is shown in Fig. 8.6. From the cycle a set of parameters can be defined that describe human locomotion:

L_c = The distance between two heel strikes of the same leg.

T_c = The time between two heel strikes of the same leg.

F_c = The cycle frequency, $F_c = \frac{1}{T_c}$.

s = The speed of the walker, $s = F_c L_c$.

Between the heel strike and the toe off the foot is in contact with the ground and so has a velocity of zero. The swing of the leg is considered to be a cyclic motion (Boulic, 1990) so a simple model for

the speed of the foot, the tip of the leg, is:

$$s_{\text{foot}}(t) = F_{\text{c}}.L_{\text{c}}\left[1 + \cos\left(2\pi F_{\text{c}}t + \phi_{\text{foot}}\right)\right] \tag{8.10}$$

where ϕ_{foot} is the initial phase of the leg in the locomotion cycle. Integrating (8.10) with respect to time gives the variation in range:

$$R_{\text{foot}}(t) = F_{\text{c}}.L_{\text{c}}\left[t + \frac{\sin\left(2\pi F_{\text{c}}t + \phi_{\text{foot}}\right)}{2\pi F_{\text{c}}}\right] + R_0 \tag{8.11}$$

where R_0 is the constant of integration and is the range at $t = 0\,\text{s}$.

The motion of the torso depends on the mode of locomotion. For walking the body has a slight forward-backward swing in addition to its gross motion, while for running the extra momentum tends to smooth out the forward-backward swing leading to:

$$s_{\text{torso}}(t) = \begin{cases} F_{\text{c}}.L_{\text{c}} + A\cos(2\pi F_{\text{c}}t + \phi_{\text{body}}) & \text{for walking} \\ F_{\text{c}}.L_{\text{c}} & \text{for running} \end{cases} \tag{8.12}$$

where ϕ_{body} is the starting phase of the body and A the amplitude of the forward-backward sway. The values of these constants have been empirically measured as $\phi_{\text{body}} = \frac{\pi}{4}$ and $A = -0.084s^2 + 0.084s$ (Boulic, 1990). Integrating this expression gives the time varying range of the body as:

$$R_{\text{torso}}(t) = \begin{cases} F_{\text{c}}.L_{\text{c}}t + \frac{A\sin(2\pi F_{\text{c}}t + \phi_{\text{body}})}{2\pi F_{\text{c}}} + R_0 & \text{walking} \\ F_{\text{c}}.L_{\text{c}}t + R_0 & \text{running} \end{cases} \tag{8.13}$$

From (8.11) and (8.13) an expression for $\boldsymbol{\mu}(t)$ was formed to complete (8.8):

$$\boldsymbol{\mu}_{\text{body part}}(t) = R_{\text{body part}}(t).\hat{\mathbf{v}} - R_0 \tag{8.14}$$

where the subscript "body part" is either "foot" or "torso". Using the equations of this section and section 4.1 it is possible to complete the terms of (8.6) for each bistatic channel of a multistatic radar.

The settings for the simulation were chosen to correspond with the planned multiperspective field trials using the UCL NetRAD system of August 2007. The NetRAD system comprises of three nodes, each capable of transmitting and receiving, allowing nine channels in the multistatic system. The precise topology of the radar, and its settings, will be described in the following section, see Fig. 8.10 and Table 8.1, but for now it will just be noted that the nodes were placed around an arc of a circle such that the target was positioned 120 m from each and the separation between the monostatic LOS directions was 24°. Additionally, the Pulse Repetition Frequency (PRF) of the radar was sufficient that the Doppler

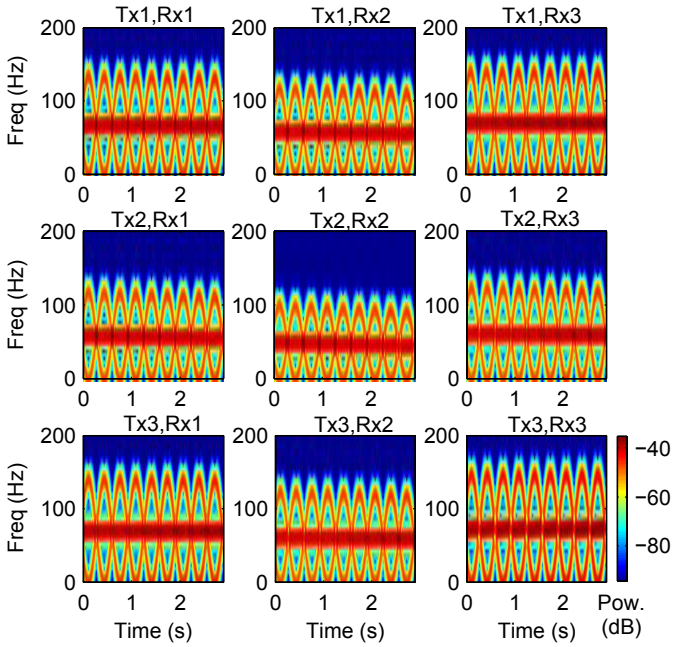


Figure 8.7: The simulated multistatic μ-DS for a person running head-on to node 3's LOS.

shifts produced by the target were well below the Nyquist frequency. For the personnel target ϕ_{foot} was set to $\pi/2$ radians for one leg and 0 radians for the other. The RCSs required to calculate the P_p^{rx} term of (8.6) were set to be: 0.7m^2 for the body and 0.15m^2 for each leg. These RCSs were an estimation based on the RCS of a person being $\approx 1\text{m}^2$ (Skolnik, 1980) and the torso being much larger than the legs. The values of F_{c} and L_{c} were set to $0.9\,\text{Hz}$ and $1.9\,\text{m}$ for the walking simulation and $1.54\,\text{Hz}$ and $3.0\,\text{m}$ for running. These values were obtained by recording the same volunteer who acted as a personnel target in the field trials walking and running a measured distance while being timed and wearing a pedometer.

The simulated spectrograms for the channels of the radar when the target was running along the node 3 LOS and walking at and angle of 135° to the node 3 LOS can be seen in seen in Figs. 8.7 and 8.8 respectively. The nine channels in the figures are ordered so that that each row represents a different node transmitting, starting with node 1, and each column represents a different node receiving, again starting with node 1. The sub-figures are labelled "Tx a, Tx b" where a is the identifier of the transmitting node and b the receiving node. The case where $a = b$ is then a monostatic channel while when $a \neq b$ the sub-figure represents a bistatic channel. The most prominant result from these simulations, compared to the monostatic models of section 4.1, is the increase in data to consider: there are now nine channels where previously there was one. However, it is apparent that there are many similarities between the channels, and that some may be redundant, but before investigating this more general features of the simulation output will be considered.

The spectrograms of Figs. 8.7 and 8.8 are as expected and match the underlying theory. At every time step each spectrogram contained three peak frequency responses corresponding to the three scat-

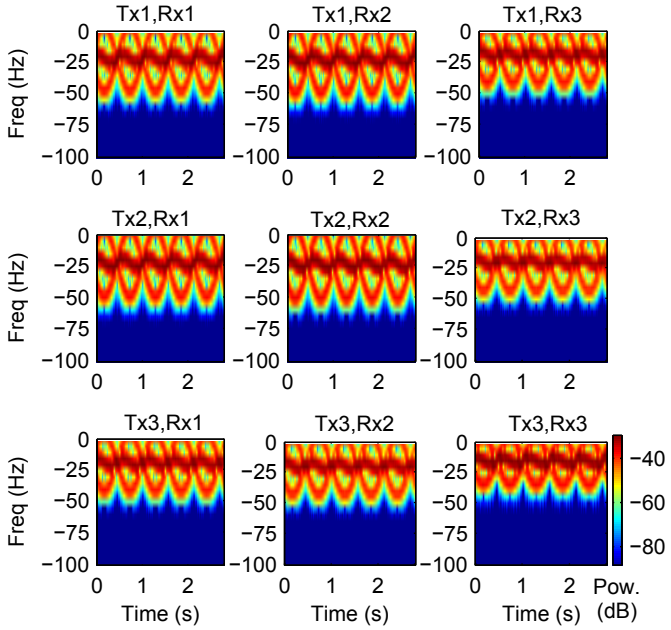
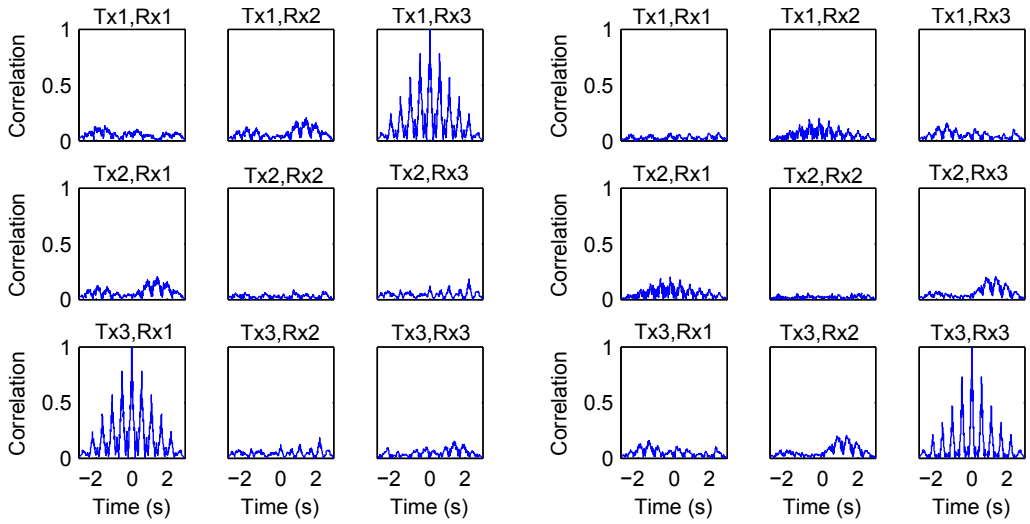


Figure 8.8: The simulated multistatic μ-DS for a person walking at 135°to node 3's LOS.

terers used in the simulation. The small amplitude responses corresponded to the limb scatterers and exhibited sinusoidal oscillations in their frequency while the variation in the response corresponding to the torso scatterer matched to mode of locomotion (running or walking). The variation, measured from the spectrograms, in frequency of the MDL and signature bandwidth—calculated from the minimum and maximum possible Doppler shifts that could be generated—were in accordance with those calculated from (8.4). It was also observed that the responses for the limb scatterers are perfectly aligned in each of the nine channels: they have the same starting phase and their peaks occur at the same time. This result was expected since the nine channels observe the target simultaneously. Each sub target (scatterer) is, therefore, in the same stage of its motion in each channel at each observation time. Combining these observations suggested that some channels of the radar were redundant, from an information perspective, since the received signals recorded by them are the same.

The information content of channels in the radar can be compared by cross-correlating them: the more similar the information of the channels the greater the correlation. Fig. 8.9 shows the correlation of two channels of the simulation and all the simulated channels; the arrangement of the sub-figures is the same as Figs. 8.7 and 8.8 and the data is from the simulation for a person walking at 135° to node 3's LOS. In Fig. 8.9 part (a) the bistatic channel node 3 transmit node 1 receive was correlated with all the other channels and it was apparent that in addition to correlating perfectly with itself, Fig. 8.9 part (a) "Tx3, Rx1", there was complete correlation with channel node 1 transmit node 3 receive, Fig. 8.9 part (a) "Tx1, Rx3". While in Fig. 8.9 part (b) the monostatic channel node 3 transmit node 3 receive was correlated with the other channels and it was apparent that the channel only correlated with itself,



(a) Node3 transmitting node 1 receiving

(b) Node 3 transmitting node 3 receiving

Figure 8.9: The correlation between the indicated channel and all of the channels for simulated μ-DS.

Fig. 8.9 part (b) "Tx3, Rx3". When this technique was applied to the other channels it was observed that all the bistatic channels correlated not only with themselves but also with their reciprocal channel while the monostatic channels correlated only with themselves. The result is significant for target classification since channels containing the duplicate information will not assist the classifier and so can be considered redundant. These results will also assist the data fusion through which the multistatic radar output is combined into a single feature vector for classification. The duplicate channels should not be included twice but used to confirm a single input representing both of them. This approach will both reduce the complexity of the feature vector and increase its accuracy. To understand why there is redundancy in the bistatic channels Fig. 8.5 and equations (8.6), and (8.7) must be considered.

The Doppler shift depends, ultimately, on the rate of change of target range (8.6) and (8.7) and this in not effected by the direction the electro-magnetic signal is travelling over the bistatic channel. If the two nodes of Fig. 8.5 are both considered to be capable of transmission and reception the Radio Frequency (RF) signal may propagate from left to right and right to left. For both propagation directions the same bistatic range of the target will be measured. The variation of this range with time will be the same for both propagation directions leading to the same baseband signal as calculated by (8.6). This rationale will apply to all the subcomponents of a target that might generate the μ-DS so the μ-DS will be the same regardless of the direction of signal propagation. (An alternative approach to this explanation is to note that (8.6) to (8.14) have no dependency on which node is transmitting and receiving, only their positions.)

With the theory of the multistatic μ-DS developed and the simulation of the multistatic μ-DS for personnel investigated the results from the field trials may be considered.

8.5 Multistatic Field Trials

University College London (UCL) has developed a three node multistatic radar system known as NetRAD (Derham et al., 2006, 2007). The system utilizes low cost RF components made available through the growth in the computer WiFi network market. Each node of the network is connected via Ethernet cables permitting distribution of the clock signal and the formation of a computer network over which commands and data may be transmitted. The use of Ethernet cabling limits the node separation to 50m. Two sets of antenna are available for the system: the first has an $8^\circ \times 8^\circ$ one way beamwidth and a 24dB gain; the second a $30^\circ \times 15^\circ$ one way beamwidth and a 15dB gain. In addition to the antennas, the receive channel offers a further 60 dB of gain. It is possible to adjust the parameters of the transmitted signal.

The radar may operate in either a CW or pulsed mode with transmit power of up to 200 mW. Transmissions are single side band. When operating in pulsed mode a variety of waveforms are available: up-chirp, down-chirp and regular rectangular pulses are possible with a maximum Bandwidth (BW) of 50 MHz. Coding can be applied to each pulse to permit truly simultaneous transmission although more common is to interleave the pulses. It is possible to operate each node at PRFs in excess of 30 kHz (the maximum PRF being dependent on the settings of other parameters) allowing interleaving rates so high that for most ground targets the pulses may be considered simultaneous. When an interleaved transmission is used each node transmits at the PRF but the transmissions of the second and third nodes are delayed such that they don't overlap with the previous node's pulse. The transmission of each node is received by all nodes of the radar.

Although the radar is able to perform matched filtering of the received signal it records a raw version of the digitized baseband signal. Each node has a data recording capacity of 512 MB and is able to export the data in binary format. After recording the exported data may be processed in Matlab using a software suite developed at UCL. The sampling of the baseband signal is at 100 MHz. The start of recording may be delayed by an integer number of samples from the transmission of a pulse to screen out direct signal feed through; a recording time is also set to limit the range recorded for pulse.

A symmetrical radar-target geometry, with the nodes 50 m apart and 120 m from the target, was used on the trial, see Fig. 8.10. Keeping the range from the node to the target area fixed resulted in the radar being positioned around the circumference of a circle. The resulting angular separation between each node's LOS was 24°. This meant that for all the bistatic channels within the system there were three possible bistatic angles: 24°, 48° and 0° (the monostatic case). The test range can have an effect on the performance of the radar due to multipath effects (Knott et al., 2004; Skolnik, 1990). Before accepting the topology shown in Fig. 8.10 it had been necessary to investigate the impact of pattern propagation factor.

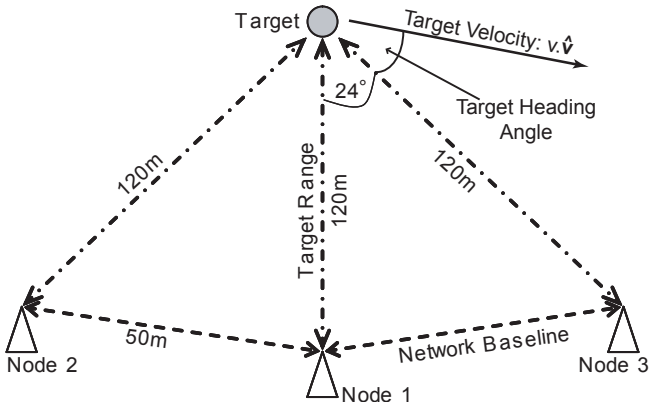


Figure 8.10: The radar-target geometry used during the field trials.

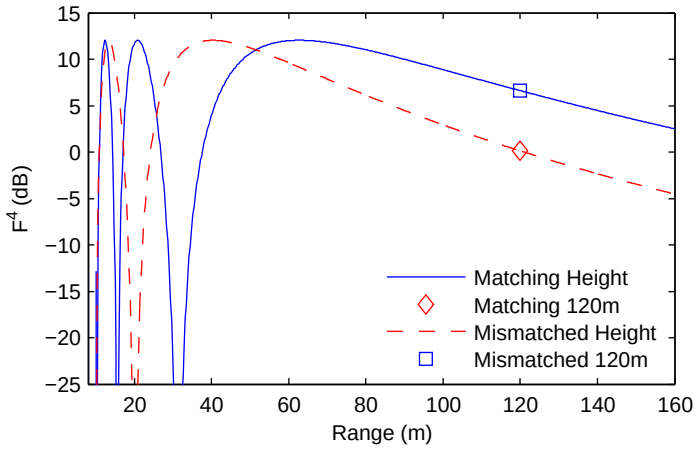


Figure 8.11: The variation of the pattern propagation factor with range for matching and mismatched antenna and target heights.

In section 2.2 the significance of the pattern propagation factor, F^4, in the radar equation, (2.3), was discussed. Due to the symmetry of the NetRAD topology during the field trials the calculation of F^4 for the bistatic channels is the same as the monostatic case since the target range from each node is the same. Fig. 8.11 shows a plot of the variation of F^4 with the range of the target from the nodes. (For production of the figure two key assumptions were made: the constant of reflection if the ground, ρ, is a perfect -1 and the ground is perfectly flat.) The blue line shows the variation of pattern propagation factor when the height of the antenna and target are matched at 1.4 m, and the dashed red line shows the case when they are mismatched with the antenna at 1.4 m and target at 0.9 m. The target height of 0.9 m was selected because it is approximately half the height of a person and it was estimated that this represented the centre of a person's RCS. In both cases the value of F^4 at 120 m is indicated and was found to be 6.6 dB for the matching height case and 0.2 dB when the heights were mismatched. Further more, the slope of the curves at 120 m is shallow so it is reasonable to assume that F^4 remains constant over the ranges that the target would move during the trial, about 3 m either side of 120 m point. This result confirmed that the selection of a 120 m target range was acceptable and that multipath would not cause problems in

Parameter	Unit	Value
Carrier Freq	GHz	2.4
Transmit Power	mW	200
Antenna Gain	dBi	21
Pulse Interleaving	–	True
PRF*	kHz	20
PW	μs	0.6
BW	MHz	40
Waveform	–	Up-chirp
Capture Time	μs	1.28
Number Of Pulses Recorded*	–	60,000
Recording Delay	Samples	181

*These values are for each node of the network.

Table 8.1: The operating parameters of the NetRAD system during the field trials.

the gathered data, such as signal cancellation. With the range confirmed all that remains to complete the trials description is the radar settings and target description.

The NetRAD operating parameters were as described in Table 8.1. The quoted PRF is the transmission PRF. Each node received at three times this rate since an interleaved transmission mode was used. For the types of target used on the trial, and because of the low velocities with which they were moving, it was considered that the interleaved PRF gave the same result as a truly simultaneous transmission.

During the trials four classes of target were used: personnel, a van, a jeep and a tractor. The three vehicle targets are shown in Fig. 8.12 and were selected based on the visibility of their wheels, that are shown in the insets of the figure. The wheels of the van are well covered by wheel arches and mud flaps and were expected to be occluded from the radar. The jeep's wheels are a little more visible since the wheel arches are large to allow off-road use and it was therefore anticipated that it may be possible to obtain some signature from this target. However, the jeep used is suitable for road use in the United Kingdom where it is illegal to have wheels that protrude beyond the chassis of the vehicle and this will limit their visibility to the radar. The final vehicle class, the tractor, had highly visible wheels due to both their size and lack of cover by the chassis. It was expected that the tractor would produce the strongest μ-DS. The personnel target used was a single individual moving in a normal manner. It was expected that a good signature would be generated by the person since their limbs were exposed and clearly visible to the radar. The motion of each target was constrained to a set of pre-defined paths and velocities.

Each target was recorded travelling along five paths that were defined relative to node 3's LOS. The target paths are shown in Fig. 8.13: the first path is at 0° to the LOS, the last 180° and the step between paths is 45°. With the exception of the tractor, targets traversed each of the paths with two different velocities: the personnel target at walking ($\approx 1.5\,\mathrm{ms}^{-1}$) and running ($\approx 4.5\,\mathrm{ms}^{-1}$) pace; the van and the jeep at 15 MPH ($6.7\,\mathrm{ms}^{-1}$) and 20 MPH ($8.9\,\mathrm{ms}^{-1}$); and the tractor traversed the paths at 15 MPH only since it could not travel at 20 MPH. The PRF and number of pulses recorded permitted 3s of data to be collected per angle-speed pairing. This time was distributed as evenly as possible either side of the

(a) The tractor target.

(b) The jeep target.

(c) The van target.

Figure 8.12: Vehicles used during multiperspective μ-DS trials. (Insets show the visibility of the wheels.)

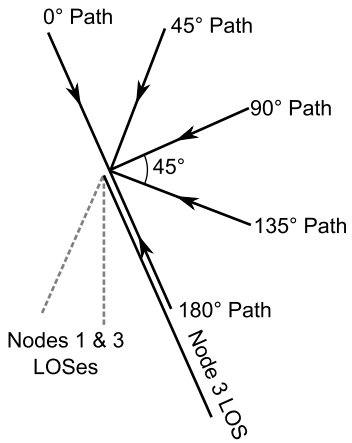


Figure 8.13: The target paths used during the multistatic radar trials.

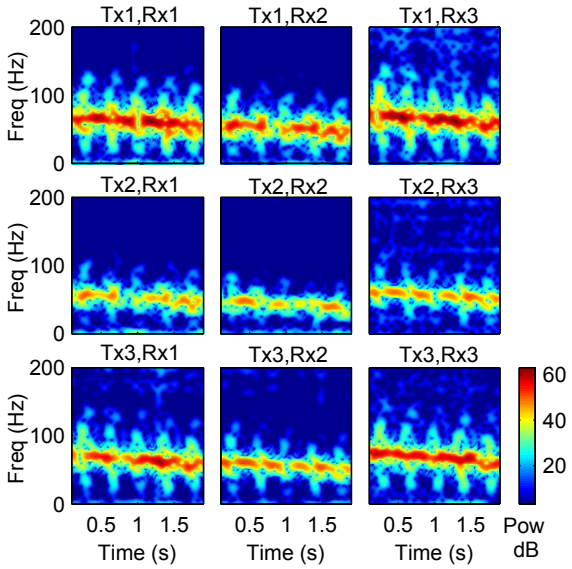


Figure 8.14: Running head-on to node 3 LOS

120 m mark. To ensure enough data was gathered every angle-speed paring was repeated five times.

The recorded data was then processed in a manner similar to the Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) data to produce a reconstructed signal per bistatic channel. The reconstructed signal was formed by manually analysing the data to see which range bins contained target returns and then coherently summing the range bins to overcome the need for target tracking. The method is described fully in section 4.4. For the NetRAD data each run of the trial result in nine reconstructed signals recorded simultaneously in time.

8.6 Analysis Of the Real Multistatic μ-DS

The trial results demonstrated different levels of μ-DS depending on the target being investigated, and the variation was attributed to the different natures of the targets. Strong signatures were measured for the personnel targets where the moving components, the limbs, were freely observable by the radar. Conversely, weak signatures were measured from the vehicle targets where the moving components were of small RCS relative to the main chassis and were often obscured by chassis. Following from this the analysis of the results will be divided into two sections: a detailed analysis of the personnel target data and a broad analysis of the vehicle data focusing on the relatively weak μ-DS observed.

8.6.1 Multistatic Personnel μ-DS

Figs. 8.14 and 8.15 show the measured multistatic μ-DS for the personnel targets. These results are directly comparable with the output of the simulation described in section 8.4 as the scenario simulated matched the field trials. In Fig. 8.14 the person is running directly towards node 3 along the node's LOS; in Fig. 8.15, the target is walking at an angle of 135° to the node's LOS. The arrangement of the sub figures is the same as for Figs. 8.7 and 8.8 with the transmit and receive nodes indicated by the sub-figure

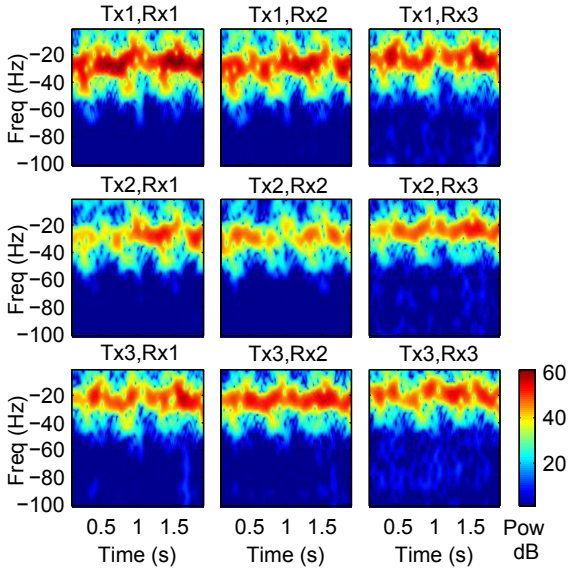


Figure 8.15: Walking 135° to node 3 LOS

titles. The intensity scale is the same for each sub figure. The trials results compared favourably with the simulation, although there were some expected differences.

The principal difference between the real and simulated spectrograms was their duration. The trial results are only given for two seconds of data, selected by removing the first and last half second of the recorded data, while the simulated results are for the full three seconds. The "pruning" of the real data was required because the antennas are not able to rotate, and so track the target, and their beam patterns did not allow a sufficient cross-range distance to illuminate the target for the entire 3 s run. As a result the initial and final pulses did not contain a target response and so were removed from the reconstructed signals created for each target. The other differences concerned the μ-DS itself.

The nature of the real signatures is much more complicated than their simulated counterparts and this was attributed to two mechanisms. The simpler mechanism concerns the distribution of the RCS over the different parts of the target. In the simulation the RCS of the torso scatterer was set to $0.7\,\mathrm{m}^2$ while the two limb scatterers were set to $0.15\,\mathrm{m}^2$; as noted in section 8.4 these numbers represent an estimate and were not expected to be correct. The real μ-DS has much lower levels for the response from the target limbs compared with the simulation. This suggested that the torso to limb RCS ratio for the simulation should have been much greater. The more complex mechanism causing a difference between the real and simulated data is that the real target is continuous while the simulated one is discrete. The backscatter radiation will be returned from the length of each limb: the tip of the swinging limb will induce a greater Doppler shift than the limb near to the joint connecting it to the torso. Furthermore, in real life limbs such as the leg do not swing with a neat, simple-harmonic motion rather they have a complicated "kicking" motion to ensure the moving foot does not accidentally strike the ground (Boulic,

Feature	Running 0° Node 3 LOS		Walking 135° Node 3 LOS	
	Simulated	Real	Simulated	Real
MDL (Hz)	75	80	-20	-20
BW (Hz)	150	120	40	40
Stepping Cycle (Hz)	1.33	1.25	0.83	0.80

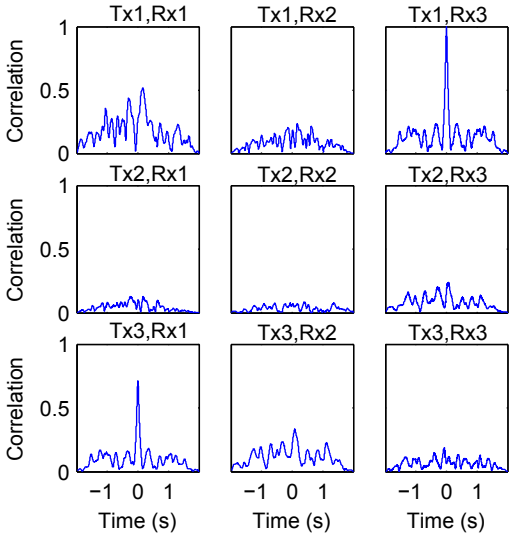
Table 8.2: Comparison of human locomotion parameters measured from node 3 Tx, node 3 Rx spectrogram.

1990). The complexity of these motions were allowed for in the model used by van Dorp (van Dorp and Groen, 2003) and at the 2008 IEEE Radar Conference an excellent paper was published allowing for both the complex motion and RCS distribution (Ram and Ling, 2008). Despite its simplicity the model used here has provided an excellent first insight into multistatic μ-DS and would still allow extraction of useful target features for radar ATR.

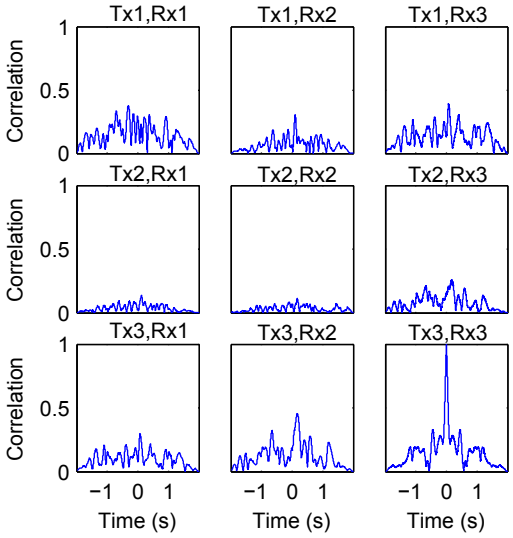
The results of the simulation and the field trials were found to be in agreement. Table 8.2 shows the values for the frequency of the MDL, bandwidth of the μ-DS and locomotion cycle frequency measured from the node 3 transmit, node 3 receive spectrogram for both the simulated and real data. It was observed that there was a good match between the real and simulated data. While this result was pleasing it was also unexpected. More sophisticated simulations have shown that the bandwidth of the μ-DS for personnel can be as large as four times the frequency of the MDL due to the kicking motion of the lower leg (van Dorp and Groen, 2003; Ghaleb et al., 2008). While this part of the target motion was not included in the simulation the target on the field trials would have exhibited the kicking action. That the high frequency components of the signature are not observed in the field trial data is attributed to the small RCS of the sub targets moving with the highest velocities. Neither van Dorp nor Ghaleb give information in their work as to the RCS they assume for the foot, which will be travelling at the peak velocity. From the spectrograms they present it is clear that the high frequency return from the foot had very low power. It is entirely possible that once noise, clutter sidelobes and the limited dynamic range of the Analogue to Digital Converter (ADC) ($84\,$dB for NetRAD) are taken into account it would not be possible to observe this component in a real situation. As such it could be considered that a simpler model will be of greater value since it will be easier to extract its parameters from measured data to form feature vectors.

The increased complexity of the real data reduced the cross-correlation between the different channels of the radar compared to the simulated case. Fig. 8.16 shows the cross-correlations for three of NetRAD's channels: part (a) and part (b) are for the same channels shown for the simulation data in section 8.4 while part (c) is for the channel node 1 transmit, node 2 receive. Again, all the examples are for a person walking at 135° to node 3's LOS.

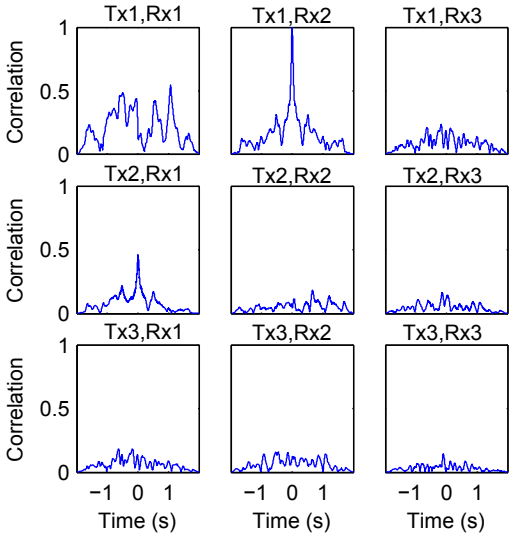
For the real μ-DS the multistatic channels still correlated with their reciprocals, but the correlations were not as strong as the simulated cases. Significant correlation was only achieved when there was



(a) Node3 transmitting node 1 receiving



(b) Node 3 transmitting node 3 receiving



(c) Node 1 transmitting node 2 receiving

Figure 8.16: The correlation between the indicated channel and all of the channels.

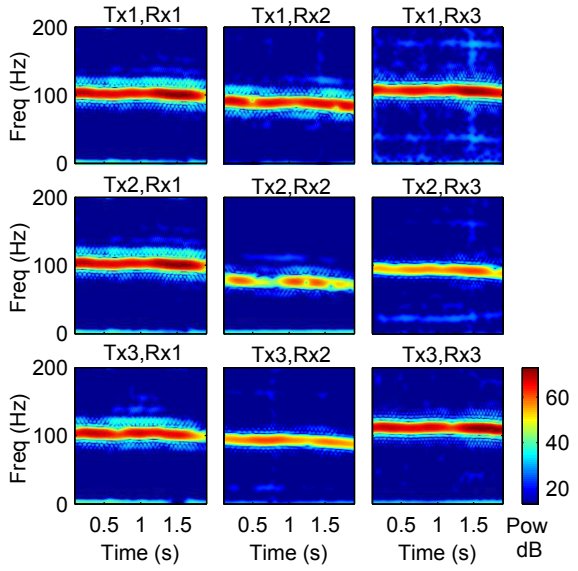
no delay between the signals with the peak failing to reach one in the reciprocal channel. The failure of amplitude of the correlation between a channel and its reciprocal to reach one was attributed to two effects. Firstly, the noise signal in the two channels was different as they use different receivers so the zero delay peak for the auto correlation case will be increased due to noise correlation. Secondly, the amplitudes of the underlying signals are different: the transmit power and receiver gains of each node of NetRAD are different due to component tolerances resulting in some channels experiencing as much as 5 dB more power than others during acceptance testing (Doughty, 2008). In Fig. 8.16 part (a) it is shown that despite the difference in amplitude the shape of the correlation curve is the same in the reciprocal channels and this suggests that the underlying target information is not unique. Again it is seen that the monostatic channels, Fig. 8.16 part (b), do not correlate with any other channels. For the simulated data the curves had multiple peaks in the channels, a result of the underlying cyclic nature of the signals. These peaks were not observed for the real signatures, a result attributed to the increased complexity of the μ-DS. With a substantially more intricate target structure, natural fluctuation in target RCS and the irregularity in the locomotion cycle of a real person the experimental μ-DS showed variations between its cycles that prevented the high amplitude peaks in the correlation curve at delays other than zero from forming. Under these more complicated conditions, the μ-DS in a bistatic channel did not always correlate with the reciprocal channel.

In Fig. 8.16 part (c) the correlation of the channel node 1 transmit, node 2 receive is shown correlated with all other channels. It was expected that the channel would correlate well with node 2 transmit, node 1 receive, see Fig. 8.16 part (c) "Tx2, Rx1", but it did not. Similar results were obtained for all cases involving node 2 transmitting. Investigation of the node 2 transmitter following the experiments determined that the lower sideband was not being correctly suppressed so its transmissions were different to the other nodes. This difference accounted for the failure of bistatic channels in which node 2 was transmitting to correlate with their reciprocals.

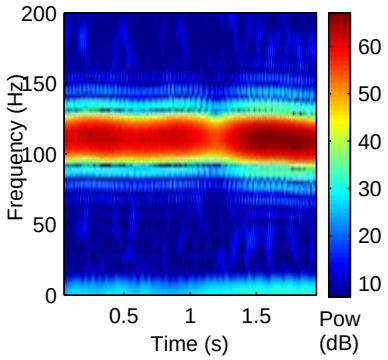
In addition to the personnel target, the μ-DS of several vehicle targets were also measured during the trial. However, the low power of NetRAD and the structure of the targets resulted in the signature being of very low power and hard to observe. Furthermore, the more complicated nature of these targets prohibited the development of a model to compare the results with. As a result the analysis of the signatures, which follows, is somewhat more qualitative than the personnel analysis, but it still provides insight into how the multistatic μ-DS may assist target classification.

8.6.2 Multistatic Vehicle μ-DS

The Thales MSTAR data had shown that the μ-DS for a wheeled vehicle was a series of spectral lines evenly distributed around the MDL. In a spectrogram these spectral lines appear as a series of horizontal stripes since they are not time varying if the vehicle has constant velocity, see section 4.4. As NetRAD



(a) Spectrograms for all channels.



(b) Spectrogram for Node 3 Tx, Node 3 Rx

Figure 8.17: Spectrograms for van target at 0° to node 3 LOS travelling at 15 MPH.

	Sidelobe					
	1st	**2nd**	**3rd**	**4th**	**5th**	**6th**
Relative amplitude Decibel (dB)	-32.2	-41.8	-48.7	-54.1	-58.5	-62.3
Distance from peak (Hz)	22.0	34.2	44.0	53.7	63.5	73.2

Table 8.3: Amplitude and distance from the MDL of of a Hanning window's frequency sidelobes.

has a lower carrier frequency and transmit power than the MSTAR it was anticipated that determining the presence, or absence, of a μ-DS in the field trials data would be difficult. The low carrier frequency would cause the spectral lines to merge together while the low power would cause their amplitude to be much closer to the noise floor. The first target to be investigated was the transit van.

The spectrograms from NetRAD's nine channels, when the van target is travelling at 0° to node 3's LOS at 15 MPH (6.7 ms^{-1}), are shown in Fig. 8.17 part (a). The arrangement of the sub plots within this figure is the same as described above. An enlargement of a single channel's spectrogram, for node 3 transmit node 3 receive, is shown in part (b). In the enlargement it is clear that there is some fine structure surrounding the MDL that could be a μ-DS or could frequency sidelobes generated by the Discrete Fourier Transform (DFT) process employed in generating the spectrogram.

During the Short Time Fourier Transform (STFT) that produces the spectrograms a 2048 point Hanning window was applied to the data during each Fourier transform; the shape of this window, and its frequency response is shown in Fig. 8.18. In creating the frequency response the signal was zero padded to 8192 samples, just as it was when creating the spectrograms, and the sample frequency has been assumed as 20 kHz to match the field trials. The frequency response has been "zoomed in" to show just the first six sidelobes, part (b) of the figure. The details of the sidelobes—their amplitude relative to, and their distance from the peak response at 0 Hz—are given in Table 8.3. To determine whether the fine

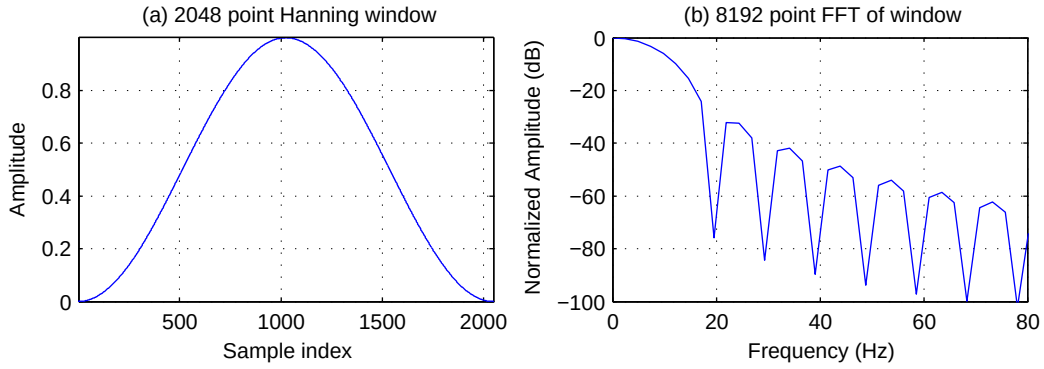
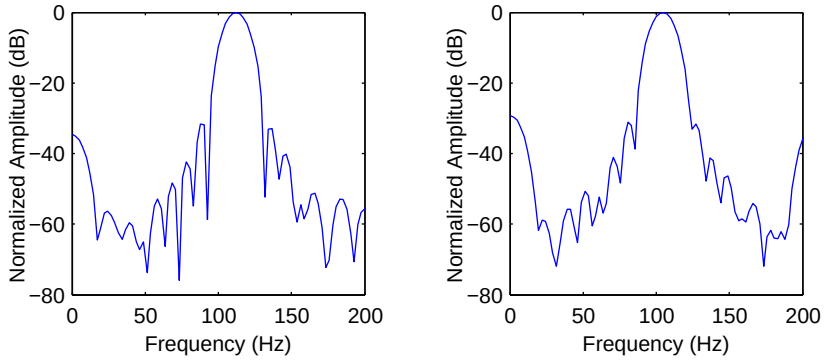


Figure 8.18: The frequency response of a 2048 point Hanning window obtained using an 8192 point FFT.



(a) Channel Node 3 Tx, Node 3 Rx (b) Channel Node 1 Tx, Node 1 Rx

Figure 8.19: Power spectra for van target at 0° to node 3 LOS travelling at 15 MPH.

structure seen in the spectrograms of Fig. 8.17 is μ-DS or frequency sidelobes it will be compared with the details of Table 8.3.

Each vertical strip of the spectrogram represents and individual Fourier transform of a short duration of the original signal. Fig. 8.19 shows the output of two of the Fourier transforms, selected at random, that go to make up the spectrograms for the channels node 3 transmit, node 3 receive and node 1 transmit, node 1 receive. These channels were chosen to provide, respectively, the case where the vehicle has the highest radial velocity, and hence Doppler shift and the case where the target aspect was closest to 45° since that angle was expected to give a strong μ-DS for the wheeled vehicles following the the Thales MSTAR data analysis. It is apparent in both parts (a) and (b) of the figure that there are low amplitude peaks to both the left and right side of the MDL. The amplitudes of these peaks, relative to MDL, and their spacing from the MDL are shown in Tables 8.4 and 8.5. It is clear that the amplitudes and spacing of the peaks are a good match with the sidelobes of Hanning window detailed in Table 8.3. It was therefore concluded that *the fine structure was a result of frequency sidelobes and was not a μ-DS.*

The next target to be considered was the tractor target that has more exposed wheels than the van and was therefore expected to produce a μ-DS. Fig. 8.20 shows the spectrograms for each channel in

	Left Peaks			Right Peaks		
	1st	2nd	3rd	1st	2nd	3rd
Relative amplitude dB	-33	-41	-47	-31	-43	-49
Distance from MDL (Hz)	24	34	44	24	34	44

Table 8.4: Frequency peak details for van target node 3 transmitting node 3 receiving.

	Left Peaks			Right Peaks		
	1st	2nd	3rd	1st	2nd	3rd
Relative amplitude dB	-35	-39	-48	-27	-46	-47
Distance from MDL (Hz)	24	34	44	22	32	44

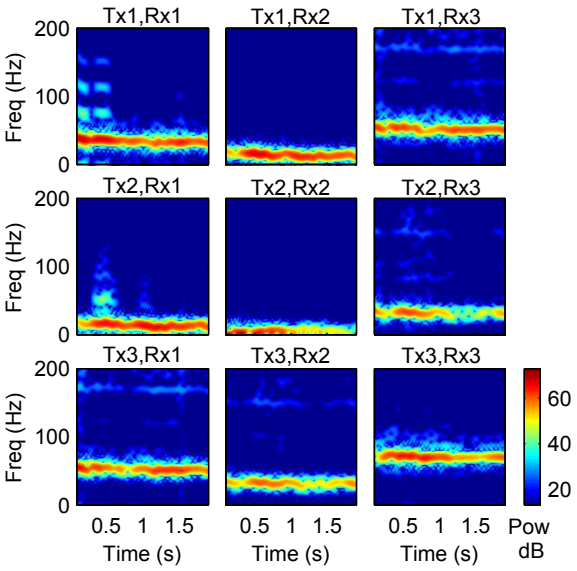
Table 8.5: Frequency peak details for van target node 1 transmitting node 1 receiving.

part (a) and a randomly selected Fourier transform from the spectrogram for channel node 3 transmit, node 2 receive in part (b). Due to the target angle and radar geometry the frequency of the MDL can be seen to vary substantially between the spectrograms. Furthermore, additional spectral lines, running horizontally across some of the spectrograms, were observed. (These lines were also observed for the jeep target and will be analysed as part of this target's consideration below.) Despite the variation in frequency of the MDL there is still fine structure visible at its edges that can be seen as peaks in the power spectra of part (b) of the figure. This time the peaks were more difficult to detect as the reduced Doppler, resulting from target aspect angle and the bistatic channel being examined, caused the signature bandwidth to narrow and the spectral lines to merge. There was only one clear peak to the left of the MDL and the first peak to the right was merged into the MDL. The details of the peaks are shown in Table 8.6 and it was again observed that the details matched with the frequency sidelobes of the Hanning window. This result was the same for all channels of the radar and it was again concluded that there was no μ-DS present in the data. The lack of μ-DS for the tractor target was unexpected. Of all the vehicle targets used, its wheels were the most exposed, see Fig. 8.12, and it had therefore been expected to generate a strong signature. To confirm the lack of signature, the trial in which the tractor was travelling at 180° to node 3's LOS was investigated.

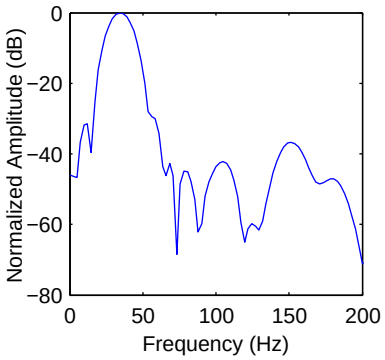
When the tractor target was moving at 180° to node 3's LOS the NetRAD channel node 3 transmit, node 3 receive was observing the rear of the target and had the best possible view of the large rear wheels. Fig. 8.21 shows the spectrograms for each channel in part (a) and a randomly selected Fourier transform from the spectrogram for channel node 3 transmit, node 3 receive in part (b). The details of the peaks surrounding the MDL are presented in Table 8.7, and again the amplitude and spacing of the

	Left Peaks			Right Peaks		
	1st	2nd	3rd	1st	2nd	3rd
Relative amplitude dB	-31	—	—	-30	-43	-45
Distance from MDL (Hz)	22	—	—	24	34	44

Table 8.6: Frequency peak details for tractor target node 3 transmitting node 2 receiving.

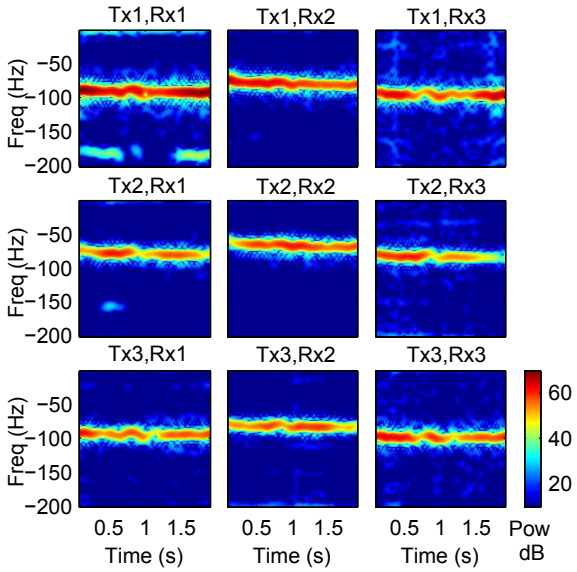


(a) Spectrograms for all channels

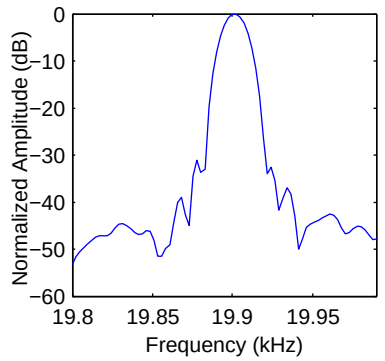


(b) Power spectrum for channel Node 3 Tx, Node 2 Rx

Figure 8.20: Tractor target at 45° to Node 3 LOS travelling at 15 MPH.



(a) Spectrograms for all channels

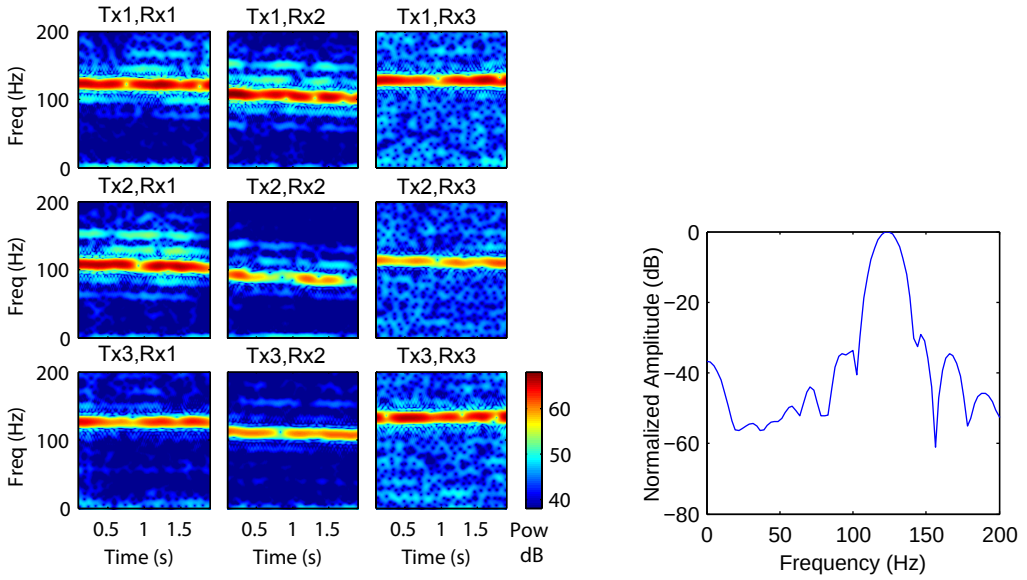


(b) Power spectrum for channel Node 3 Tx, Node 3 Rx

Figure 8.21: Tractor target at 180° to Node 3 LOS travelling at 15 MPH.

	Left Peaks			Right Peaks		
	1st	2nd	3rd	1st	2nd	3rd
Relative amplitude dB	-31	39	—	-33	-37	—
Distance from MDL (Hz)	24	34	—	22	32	—

Table 8.7: Frequency peak details for tractor target travelling at 180° to node 3 LOS, node 3 transmitting node 3 receiving.



(a) Spectrograms for all channels

(b) Power spectrum for channel Node 1 Tx, Node 1 Rx

Figure 8.22: Jeep target at 0° to Node 3 LOS travelling at 20 MPH.

peaks coincided with the position of the frequency sidelobes for the Hanning window. Once again it had to be concluded that there was no μ-DS visible in the tractor data.

The final target to be analysed was the jeep target. Fig. 8.22 shows the results for the trial when the jeep was travelling at 20 MPH and at 0° to node 3's LOS. The example power spectra shown in part (b) was selected from the spectrogram for the node 1 transmit, node 1 receive channel. As for the other targets the amplitude and spacing of the three peaks either side of the MDL are detailed in Table 8.8. The match between the peak details and the frequency sidelobes of the window was not as strong for the jeep as it was for the other target. Based on the spacing alone, it appeared that to the left of the MDL the 1st, 2nd and 4th sidelobes were present, but not the 3rd; to the right, the 1st, 3rd and 5th appeared present. Furthermore, the amplitudes of several of the peaks were larger than the amplitudes of the frequency sidelobes for the window. This was particularly evident to the right of the MDL in spectrum of Fig. 8.22

	Left Peaks			Right Peaks		
	1st	2nd	3rd	1st	2nd	3rd
Relative amplitude dB	-34	-34	-44	-29	-35	-46
Distance from MDL (Hz)	24	32	54	22	42	66

Table 8.8: Frequency peak details for jeep target node 1 transmitting node 1 receiving.

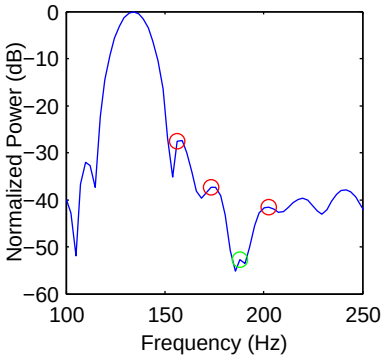


Figure 8.23: Power spectrum for the jeep target from channel node 3 transmit, node 3 receive.

part (b).

The reduced match between the peaks surrounding the MDL and the window frequency sidelobes for the jeep target suggests there is a small μ-DS present in the data. Rather than manifest itself as a series of spectral lines, as had been expected, the μ-DS acted to increase the power of the target's frequency response in the region surrounding the MDL. This power increase was particularly significant in the frequencies above the MDL. Since the high frequency responses would correspond to the signal backscattered from the top of the wheel, it was speculated that the increased power at these frequencies was a result of the cavity formed between the top of the wheel and the chassis, see inset of Fig. 8.12 part (b), acting as a corner reflector with large RCS. The effect was particularly evident in the channel node 3 transmit, node 3 receive. An example power spectrum, taken from the spectrogram for this channel, is shown in Fig. 8.23. The first four peaks to the right of the MDL have been highlighted. All the peaks are distances from the MDL that correspond to frequency sidelobes of the window function. However, it is only the peak highlighted in green that has a power comparable to the sidelobe at that distance; those peaks highlighted in red have powers ranging from approximately 5 to 10 dB above those for the appropriate window sidelobes. Although these results suggest there is a μ-DS for the jeep target it remains unclear how useful features for classification could be extracted from it since it has the form of a slight increase in power over a region of frequency surrounding the MDL.

The spectrograms in Figs. 8.17 to 8.21 show a strong MDL and several additional lines at lower power; the last part of the analysis for the wheeled vehicle field trials data is to consider these extra lines. Like the MDL these lines indicate a time stationary return from the target. Unlike the MDL the lines are at a low power and were observed to have time fluctuations in their power amplitudes to an extent where, at times, the amplitude fell below the noise floor. When the range of the y-axis was increased it was noted that the lines appear throughout the frequency space. Following from this it was concluded that the lines were not part of a μ-DS that might accompany the target. The maximum frequency of a μ-DS component for a wheeled target is twice that of the MDL and corresponds to the velocity of the top of

the wheel that is twice the chassis velocity. However, the frequency of the lines does vary between the different channels of the radar and it was observed that their spacing from the MDL was the same in all channels. As a result, the lines were attributed to out of band transmissions from NetRAD. NetRAD's transmission performance is reasonable with no high-power out of band transmissions (Derham et al., 2007; Doughty, 2008), however, it is known that the oscillators used do produce extra spur transmissions but at low powers. The use of a 2048 point DFT results in an integration gain of 66 dB and this would be sufficient to allow the reflection of the spur transmission from the target to be seen.

Overall the results for the vehicle part of the field trials were disappointing. Unlike the personnel part of the trials there was no clear μ-DS visible in the spectrograms generated from the data. Detailed analysis of the fine structure that was visible in the spectrograms revealed it to be a result of the frequency sidelobes of the window function used rather than a μ-DS. The only vehicle target to exhibit any indication of a signature was the jeep target. The power of the frequency sidelobes above the MDL was higher than predicted by theory, a result attributed to the cavity formed between the top of the wheel and the main chassis acting as a high RCS reflector. The failure to detect any meaningful μ-DS in the vehicle data limited the multistatic μ-DS classification that could be attempted for vehicle targets.

8.7 Multistatic Micro-Doppler Signature Classification

The dataset used for multiperspective μ-DS classification was different to that described in sections 8.5 and 8.6 as it was gathered at a separate trial. Although the trials are very similar the dataset that was used for the classification tests is not of as high a quality at that analysed above. As will be discussed NetRAD's transmit power was set to just 10 mW and fewer pulses were recorded. These limitations at the field trial resulted in a dataset in which the μ-DS of the targets were hard to observe. The reason for this reduction in quality is that the dataset used in testing was gathered before that analysed in section 8.6 and its failings used to plan the subsequent trial. Unfortunately there was insufficient time to undertake classification testing with the newer data and the results presented here are for the original dataset that contained limited examples of the μ-DS. During the remainder of this section all references to datasets and trials will be for the original trial that provided the data used in the classification tests.

The topology of NetRAD during the field trials was as shown in Fig. 8.10. The nodes were separated by 50 m and placed around an arc to give symmetrical views of the targets; the target position was 120 m from each node. The antennas do not rotate so capture time was limited by how long it takes the target to traverse the multistatic region where the antenna main lobes overlap. One or two seconds of data was gathered depending on the target being used. The settings of the system parameters are given in Table 8.9. As in section 8.5 an interleaved pulse mode is used, but given the high PRF and the velocities of the targets used the pulse were considered as having simultaneous transmission.

Three classes of target were used: a wheeled vehicle, personnel and a person cycling. The personnel

Parameter	Unit	Value
Carrier Freq	GHz	2.4
Transmit Power	mW	10
Antenna Gain	dBi	21
Pulse Interleaving	–	True
PRF*	kHz	10
PW	μs	0.6
BW	MHz	40
Waveform	–	Up-chirp
Capture Time	μs	2.56
Number Of Pulses Recorded*	–	10×10^3 or 20×10^3
Recording Delay	Samples	181

*These values are for each node of the network.

Table 8.9: NetRAD operating parameters when gathering classification data.

and bicycle targets were recorded travelling head-on to, tail-on to and perpendicular to node 3's LOS while the wheeled vehicle was only recorded at head-on and perpendicular aspects. The person moved at walking pace, the cyclist attempted to go as quickly as conditions allowed and the vehicle speed was kept as close to 15 MPH ($6.7\,\text{ms}^{-1}$) as possible. During the wheeled vehicle trials 10×10^3 pulses were recorded, giving 1 s of data, while 20×10^3 pulses were recorded for personnel and bicycle, giving 2 s of data. The lower number of pulses for the wheeled vehicle was a result of the speed with which it traversed the multistatic region.

The wheeled vehicle used during the trial was a Ford Transit Van, the same as the one shown in Fig. 8.12 part (c). When the trial was planned it was hoped that it would give a similar signature to the wheeled vehicle target from the Thales MSTAR dataset, subject to the different carrier frequency. The personnel target was also expected to give a signature comparable to the MSTAR data. There was no comparable target for the bicycle in the Thales data, but it was anticipated that its signature would be similar to that of personnel since its μ-DS would also be generated be moving limbs. It was not thought that the wheels of the bicycle would contribute to the μ-DS as, by design, they are not comprised of much solid material so their RCS would be small.

An example spectrogram for the bicycle target is provided in Fig. 8.24. It is taken from the channel node 3 transmit, node 3 receive when the bicycle was travelling along node 3's LOS. As expected the result is similar to that of personnel. There is a MDL with a constant frequency, in this instance of $\approx$ 80 Hz, with extra signature components surrounding it that show oscillations in frequency. These components are attributed to the legs of the cyclist. It was observed that the lower frequency of the signature is not 0 Hz confirming that the wheels can not be detected by the radar. The bottom of the wheels were the only parts of the bike in contact with the ground and therefore had zero velocity and could have generated a frequency component at 0 Hz. It may be thought that at the bottom of the peddling action the foot of a cyclist has zero velocity, and even starts to move backward, however, *this is in the reference frame of the bicycle only*. In the radar's frame of reference the foot would still be moving

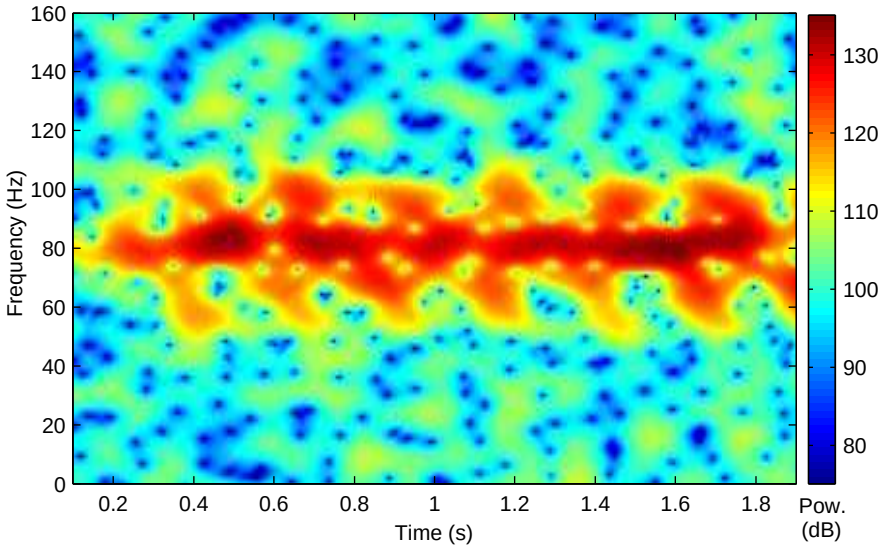


Figure 8.24: Spectrogram for a bicycle travelling towards node 3's LOS.

forward, albeit more slowly than the rest of target, due to the gross motion of the bicycle. The minimum frequency of the μ-DS would therefore be greater than zero, as was observed in Fig. 8.24. The analysis of the data from the other targets gave results comparable to those of section 8.6 but with lower power due to the lower radar transmission power, see Tables 8.1 and 8.9.

The classification method employed was DTW since it is able to cope with multi dimensional data. However, different pre-processing was required for the multiperspective data. Since the nine channels of the NetRAD system needed to be merged in to a single feature vector.

8.7.1 Pre-Processing The Multistatic Data

The pre processing of the NetRAD data was performed manually. It was five step activity: removal of zero Doppler clutter; selection of the range bins containing the target; formation of a reconstructed signal for each bistatic path; segmentation of the reconstructed signals into short frames; and creation of reference and test datasets.

Using $f_{\mathrm{D}} = 2v_{\mathrm{r}}\lambda_{\mathrm{c}}^{-1}$, where v_{r} is the radial target velocity and $\lambda_{\mathrm{c}} = 0.125\,\mathrm{m}$ is the carrier wavelength, resulted in a predicted maximum Doppler shift of $\approx 20\,\mathrm{Hz}$ for personnel. This is close in frequency to stationary clutter so the zero-Doppler clutter was removed to ensure the target signals were not dwarfed by clutter sidelobes. The method used was different from that of section 5.4. Because the NetRAD data had more range samples than the Thales MSTAR data it was possible to use a common routine from ISAR processing to remove the zero-Doppler clutter (Showman et al., 1998).

Production of a reconstructed signal for each bistatic channel was performed in the same manner as for the Thales MSTAR data. The range bins containing the target response were manually identified and coherently summed. This provided a set of nine signals (one for each channel) per target, sampled at the

PRF. However, the division of the reconstructed signals into short frames to be used in classification and their subsequent distribution into the reference and test datasets was different to the approach taken for the Thales data.

Each data frame needed to contain at least 3×10^3 samples to allow the μ-DS to be observed. This number had been obtained empirically during the analysis of the data from the trial by manual observation of power spectra calculated from different numbers of samples. It had been observed that in order to see a μ-DS for the personnel and bicycle targets, at least 3×10^3 samples needed to be included in each DFT of the spectrogram. This high number of samples gave an integration gain of $\approx 70\,\text{dB}$ and helped counter the low transmit power of the radar compared with the data analysed in section 8.6. However, when the frame size was set to 3×10^3 samples it transpired that there was insufficient computer memory available to calculate the local distance maps of the DTW algorithm. As a result the frame duration had to be reduced to 1×10^3 samples per frame.

With so many samples used per data frame, it was necessary to overlap the frames to generate sufficient data to form a reference and test dataset from the trials data. For the van target there were only 10×10^3 in each reconstructed signal and for the personnel and bicycle targets 20×10^3 samples. Without overlapping this would have lead to 10 and 20 data frames respectively. It was decided that fifty frames per target would be generated. Twenty were randomly selected for use as a reference dataset, and the remaining thirty were used for testing. It is important to note that the same twenty frames were selected from the reconstructed signal for each of NetRAD's channels. To produce this number of data frames the frames were overlapped by 82% for the wheeled vehicle and 61% for personnel and the bicycle. The reference dataset contained examples from the target travelling head-on to node 3 only, while the test dataset used frames from available all target aspects. The datasets created were different from those used with the Thales MSTAR data, defined in section 5.3, since they contained examples from multiple channels. Part of the multistatic DTW process would be to fuse the information from these channels together.

8.7.2 Multiperspective Classification

The corresponding data frames from NetRAD's multiple channels were fused together by making each sample of the feature vector a vector in its own right. The feature vectors that would be used by the radar then became data series where each sample of the series was a nine element vector; each element of these vectors was a sample from one of the radar channels. Under this scheme a data series, $U(m)$,

would be mathematically described by:

$$U[m] = \left\{ \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{1,N} \end{bmatrix}, \begin{bmatrix} a_{2,1} \\ a_{2,2} \\ \vdots \\ a_{2,N} \end{bmatrix} \dots \begin{bmatrix} a_{M,1} \\ a_{M,2} \\ \vdots \\ a_{M,N} \end{bmatrix} \right\} = \{\mathbf{a}[1], \mathbf{a}[2] \dots \mathbf{a}[M]\} = \{\mathbf{a}[m]\} \tag{8.15}$$

where $\mathbf{a}[m]$ is a vector-sample containing information from all of the nine channels, i.e. $N = 9$; $a_{m,n}$ is the m^{th} individual sample from the n^{th} channel; and M is the number of vector-samples in the series. From the pre-processing described above $M = 1 \times 10^3$. Although these data series seem very different from the feature vectors used in the original DTW testing of Chapter 6 they actually conform to the description of the DTW algorithm provided in section 6.1.

The original description of DTW provided described how it compared two time series $T[j] = \{\mathbf{t}[j]\}$ and $R[k] = \{\mathbf{r}[k]\}$. It is important to note that the series described by (6.1) and (6.2) have vectors at each of their samples, just as $U[m]$ does. The local distance calculated at each sample of the time series in the original definition was the Euclidean distance that can be calculated as easily for vectors as it can for scalars. It is therefore possible to use vector-series, such as $U[m]$, in the DTW algorithm as it stands without any modification. The only considerations needed are that memory and number of processor operations required to calculate the normalized global distance have increased due to the extra channels. It was this increase in computational demand that resulted in the length of the data series being limited to 1×10^3, rather than the ideal 3×10^3, samples.

The testing method for the multiperspective μ-DS classifier was much simpler than its monoperspective counterpart. Due to the lack of available test data there was only a single frame duration, of 100 ms, being tested. There was no use of different correlation levels since the low number of frames had required a random selection of entries for the reference dataset. With such simplistic datasets it was not possible to test for unknown class detection so the "unknown" threshold of the classifier was disabled. The two test statistics used to analyse the data were the probability of correct classification and the probability of generalization. It was possible to consider the latter of these as there were additional target aspects in the test dataset compared to the reference.

8.7.3 Multiperspective Classification Results

Table 8.10 shows the confusion matrix for the multiperspective tests. The P_{cc} obtained is an impressive 0.97 while P_{Gen} is just 0.24. The poor performance for P_{Gen} was attributed to the classifier working on velocity alone and not the μ-DS. The 1×10^3 pulse frames were identified above as being too short to observe the μ-DS so it is unlikely this information was used by the classifier. Instead target velocity has been used. Each target class has its own velocity: the wheeled vehicle fastest, personnel slowest

	Declared Class		
Input Class	**Wheeled Vehicle**	**Bicycle**	**Personnel**
Head-on Wheeled Vehicle	1.0	0.0	0.0
Perpendicular Wheeled Vehicle	0.0	0.2	0.8
Head-on Bicycle	0.1	0.9	0.0
Perpendicular Bicycle	0.0	0.0	0.0
Tail-on Bicycle	0.0	0.5	0.5
Head-on Personnel	0.0	0.0	1.0
Perpendicular Personnel	0.0	0.0	1.0
Tail-on Personnel	0.0	0.0	1.0

Table 8.10: Confusion matrix for the multiperspective μ-DS classifier.

and the bicycle in between. In the head-on cases the velocities are sufficiently different for the classifier to successfully identify the targets. However, when the target motion is perpendicular to node 3's LOS the bistatic component of the velocity in each channel is reduced resulting lower Doppler shifts. These lower Doppler shifts are then most similar to the personnel target entries in the reference dataset so the classifier declares the classes as personnel.

Although this testing is significant as it represents the first attempt in the open literature to perform multiperspective μ-DS classification (Smith et al., 2007) it was a very limited experiment. The principal limitation is the heavy overlapping of the data frames that was required to generate enough frames to conduct a test. It is no surprise that the classifier obtains such a high P_{cc} when it the test inputs are 82% or 61%, depending on target, the same as the frames of the reference dataset. Following this, as is discussed above, the low power setting of the radar during the trial meant that the presence of a meaningful Doppler signature in the data series used in the testing is questionable. The final significant limitation relates to the data fusion method employed.

By fusing the channels together into a single multidimensional data series the reference dataset becomes specific to the radar-target geometry used during the trial. The nine channels available from NetRAD were combined to form a nine dimensional feature space and each data series then became a trajectory through that space (Sankoff and Kruskal, 1999) i.e. each sample represents a point in the space and the location of the point changes at each sample. To ensure this process worked, the ordering of the channels in each data series created was kept the same. The DTW algorithm is then comparing trajectories rather than scalar time series as it was in the monoperspective case. If a different radar-target geometry were used, or the ordering of the channels changed, then the relationships between the elements of a vector in the space would change and a new reference dataset would be required. Since target location affects the radar-target geometry this is a severe limitation on this approach. Alternative methods for fusing multiple channels together have been proposed in the literature that do not suffer this problems. Of particular note is a method of probability multiplication that has been used in HRRP classification (Vespe et al., 2007) and the use of HMMs (Robinson et al., 2005). Unfortunately, neither

of these would have been suitable for use with the DTW classifier.

As a result of the limitations it is not reasonable to draw firm conclusions from the multiperspective μ-DS classification testing. The testing has shown that time domain based DTW can process multiperspective data although the simplistic data fusion method taken is unlikely to succeed under more robust testing. The work that has been described in this chapter has shown that there is relevance in considering the multiperspective μ-DS for target classification. The theory and analysis of the earlier sections showed that multistatic data would be robust to some of the difficulties faced in automatic target recognition. At its most basic, multiperspective approaches overcome self occlusion since it is highly unlikely all aspects will suffer occlusion. More important than this is the extra information a multiperspective approach provides. The μ-DS of a particular perspective gives information about motion of sub-components along that perspective: provided the perspectives are suitably spaced each one is therefore capable of providing unique information about the target. Such approaches could lead to a rapid increase in the amount of information that must be processed per classification. Fortunately, the analysis of the simulation and field trials data demonstrated that redundancy in the data maybe exploited to reduce the amount of data to process when working with multiple perspectives. It is unfortunate that there was insufficient time for the high quality dataset gathered to be used in classification testing and that an earlier dataset containing limited examples of the μ-DS was relied on. The method by which further multiperspective μ-DS testing could be conducted will be discussed in the *Future Work* section of Chapter 9.

8.8 Multistatic Micro-Doppler Review

The theory and classification activities of this chapter summarize the first reported investigations of multiperspective/multistatic μ-DS and its classification in the open literature. The work has previously been reported in two conference papers: the first at the 2007 IET International Radar Conference held in Edinburgh (Smith et al., 2007) and the second at the 2008 IEEE Radar Conference held in Rome (Smith et al., 2008b). Prior to these publications the only non-monostatic μ-DS work that had been conducted was an investigation of the μ-DS of a helicopter illuminated by a bistatic radar (Johnsen et al., 2003). While this work does provide valuable observations on the nature of a helicopter μ-DS the radar used is still single channel (single perspective) and classification is not attempted. As such, despite the limited nature of the μ-DS for some of the targets the work presented here is a significant advance in this novel target recognition technique.

The limited μ-DS observed in the field trials data is primarily a result of the operating parameters of the NetRAD system and the nature of the targets used. The basic concept of multistatic μ-DS as an ATR tool is still valid. NetRAD was not designed to perform micro-Doppler investigations and the failure to detect significant micro-Doppler responses for the wheeled vehicle targets results from the low carrier frequency of $2.4\,\text{GHz}$ and the low transmit power of $200\,\text{mW}$. In comparison, the Thales MSTAR radar,

that collected the monostatic μ-DS data, that has a transmit frequency in the Ku-band and a transmit power of $4\,\mathrm{W}$ making it much more likely to detect the μ-DS. Unless the components of the target that generate the signature are relatively large and exposed to the incident RF radiation it is unlikely the μ-DS would be observed by NetRAD. This is why the personnel and bicycle targets, where the moving components are limbs that constitute a high percentage of the target RCS, result in a detectable μ-DS and the vehicle targets, where the moving components are small and occluded from the radar, do not. However, the monostatic results, see section 4.4, demonstrate that these targets do have a μ-DS and it can be classified (Stove and Sykes, 2003; Jahangir et al., 2003b,a; Smith et al., 2006b,a,c, 2008c), see Chapters 6 and 7.

If the transmitters on the multistatic radar used to gather the data had been comparable to the Thales MSTAR then it is reasonable to hypothesise that the μ-DS for the vehicle targets would have been detected. Wheeled vehicles do exhibit a μ-DS, see section 4.4, and the multistatic results for personnel confirm that monostatic signatures readily convert into multistatic ones and are detectable with a multistatic radar. With a dataset collected by such a system it is expected that the results of radar ATR testing would have been much different. Results for HRRP data based radar ATR of air targets have shown gains in the P_{cc} of between 0.25 and 0.30 when the method by which multiple perspectives are fused was changed from a simple averaging, that reduced to problem to a monoperspective one, to the use of HMMs that classified the multiperspective data directly (Zhu et al., 2007). HMMs have also been shown to improve the discrimination rate between ground targets and trees in Synthetic Aperture Radar (SAR) images (Runkle et al., 2000). Depending on the false alarm rate that is acceptable the HMMs can result in improvements of up to 20%. Furthermore, the classification of ground vehicles has been shown to improve dramatically when multiple aspects are presented to the classifier (Vespe et al., 2007). Vespe's work tests three HRRP data classifiers–k-Nearest Neighbour (k-NN), naïve Bayesian and feed-forward ANN–all of which are modified to process multiperspective data. Only results for P_{cc} are provided, however, gains of up to 0.3 are possible when five perspectives are used. Using Vespes variant of the naïve Bayesian classifier with the μ-DS feature extraction method presented in section 5.4 would therefore be expected to give impressive performance.

Overall, the potential of multiperspective μ-DS based radar ATR appears promising. It is clear from the results presented in this chapter and the analysis of section 4.4 that given suitable equipment a multistatic μ-DS could be detected and this could provide the input for a multiperspective classifier. The results presented by Vespe for the modified naïve Bayesian classifier (Vespe et al., 2007) suggest that the high performance of a regular naïve Bayesian classifier working with frequency domain μ-DS, see Chapter 7, could be improved even further. With such high potential it is clear that multiperspective μ-DS has significant potential as a future radar ATR technique.

8.9 Summary

The investigation of multistatic μ-DS and its value in radar ATR was broken into three broad sections: the extension of existing μ-DS theory into the multistatic arena; the gathering and analysis of multistatic μ-DS data; and an attempt to perform multiperspective μ-DS based radar ATR.

Following a review of the concepts of multistatic radar, a theory of the multistatic μ-DS was developed by replacing the expressions for monostatic range and phase from the theory of Chapter 4 with their bistatic counterparts. It was observed that a multistatic radar resulted in several data channels and that the μ-DS signal had to be calculated in each of these. A model for a simplified personnel target was developed to investigate the multistatic μ-DS. It was seen that for a three node radar, where each node could both transmit and receive, there were nine channels, each of which containing a μ-DS. Similarities between the channels were observed, resulting from each one providing a simultaneous observation of the target, but from a different aspect. A correlation based analysis of the simulation data showed that there was strong cross-correlation between the μ-DS in reciprocal bistatic channels. This reciprocity lead to a redundancy in the information provided about a target by a multistatic radar. However, the redundancy was also considered a potential benefit since the number of channels increased as the square of the number of nodes in the radar leading to a potential data fusion problem. Being able to ignore some channels as redundant could therefore assist a radar ATR solution by reducing the amount of data to be fused. With the theory developed, the field trial data was analysed.

The collection of a dataset containing four targets—personnel, a van, a jeep and a tractor—using UCL's NetRAD system was described before the data was analysed. The first set of target data to be analysed was for the personnel target. The parameters of the simulation had deliberately been selected to be the same as the field trial and it was observed that, broadly, the simulation output matched the real data. There was more detail in the real data, however. This result was expected result since the simulation target was simplified, but despite the simplification it was apparent that the main μ-DS features were the same. The cross-correlation analysis was repeated using the real personnel target data and it was again seen that reciprocal bistatic channels correlated and so could be considered redundant. Unfortunately, the results for the vehicle targets indicated that only limited μ-DS features had been detected. By comparing the features visible in the spectrograms for the wheeled vehicle targets with the Fourier transform of the window function it was determined that the only vehicle target to exhibit a μ-DS was the jeep. Even then the μ-DS took the form of a slight increase in power over a small frequency bandwidth surrounding the MDL. It was unclear how this signature could be used for classification since it did not display any appreciable features. The failure to detect target signatures was attributed the the NetRAD system having a low carrier frequency and transmit power and the vehicle targets' chassis acting to occlude the moving components that would have generated a μ-DS.

The attempt to perform multiperspective radar ATR using multistatic μ-DS data was limited due to the lack of significant signatures detected for vehicle targets. The classification testing was conducted using a different dataset to that used during the signature analysis work. In the classification dataset the targets were: personnel, a bicycle and a van. An example of the μ-DS associated with a bicycle showed it to be clearly detectable and different to that associated with personnel. The size of the dataset was limited, so that when the data was divided into frames they had to be overlapped. Reference and test dataset frames were chosen by random selection rather than the correlation method. The data fusion requirement for the classifier, originating from the multistatic radar producing several channels of data, was met through use of the DTW algorithm. DTW is capable of processing data series where each sample is vector of data: by suitably arranging the data from each of NetRAD's channels a classifiable data series was produced. The performance of the classifier was analysed in terms of probability of correct classification, P_{cc}, and generalization, P_{Gen}, only. While the P_{cc} result looked impressive the P_{Gen} result was poor. More detailed investigation showed that the high P_{cc} was a result of the overlapping of the data frames rather than the classifier working especially well. This disappointing result was in contrast to the use of multiperspective classification in other radar ATR applications. To close the chapter the current literature in this area was reviewed.

The review of multiperspective radar ATR showed that for HRRP and SAR based systems the fusing of multiple target aspects had resulted in performance improvements. Of particular interest was the use of a naïve Baysian classifier that had shown particularly impressive performance gains when modified to work with multiperspective data. As a result of the review it was postulated that the poor results for multiperspective classification seen in this research were a result of the data, and the limited μ-DS it contained, rather than the general methodology. It was hypothesized that if a suitable dataset could be obtained then, given the high performance of the naïve Baysian classifier in Chapter 7, the multistatic version would give excellent result for μ-DS based radar ATR.

Chapter 9

Conclusions And Future Work

To close this thesis a set of definite conclusions are drawn from the novel research described in the previous chapters. The chapter closes with the identification of a series of areas for further investigation that future researchers may wish to consider.

9.1 Conclusions

The principal contribution of the research presented in Chapters 2 to 8 was described in section 1.2 and the summaries at the end of each chapter. From these contributions it is possible to draw several conclusions.

To completely evaluate a radar Automatic Target Recognition (ATR) system more evaluation parameters than the probability of correct classification, P_{cc}, are required. The classifier's reliability, RL, probability of generalization, P_{Gen}, should also be evaluated. These parameters were introduced in Chapter 3 and describe, respectively, the probability the the input class is correctly declared (P_{cc}), the probability that the input class was the same as the declared class (RL) and the ability of the classifier to extrapolate, or generalize, from its reference data. The application of the parameters to the results in Chapters 6 and 7 demonstrate the improved understanding of the classifiers they provide.

To be suitable for use in operational situations, a radar ATR solution must be capable of declaring inputs as being from an unknown target class. An unknown target class is a one not represented in the classifier's reference dataset. The effectiveness with which a classifier detects and declares "unknowns" is measured with two further parameters. The probability of declaration, P_{Dec}, measures the probability a known class has of passing the "unknown" threshold while the probability of false alarm, P_{FA}, measured the chance an unknown class of being incorrectly declared as one of the knowns.

Receiver Operating Characteristic (ROC) curve analysis provides an excellent method for evaluating the correct setting of an "unknown" detection threshold. In Chapter 3 it is proposed that thresholding the internal scores of a classifier will allow unknown target inputs to be detected. Changing the level of the threshold affects the values of P_{Dec} and P_{FA} and the two may be used to produce ROC curves

allowing the optimum threhsold level to be found.

To understand if a given classifier is performing well, i.e. its values of P_{cc} etc. are acceptable, it is necessary to estimate the best possible performance of any classifier from the reference dataset. The classifier's performance can then be contrasted with the estimate. The method by which the estimation of the evaluation metrics proposed in the research may be achieved is not straightforward. A method for predicting P_{cc} is proposed in Chapter 3 based on the Bhattacharyya bound on the error. The method is applied to real data in Chapter 7 and demonstrated as being successful.

The micro-Doppler effect, and the resulting target Micro-Doppler Signature (μ-DS), is a suitable phenomena on which to base a radar ATR solution. This represents a key conclusion of the research and is supported by the entire thesis and the publications available in the open literature. Following from this broad conclusion are a series of smaller conclusions concerning specific methods by which the Micro-Doppler Signature (μ-DS) may be classified.

The Dynamic Time Warping (DTW) classifier is capable of discriminating between three common classes of radar ground target–personnel, wheeled vehicles and tracked vehicles–based on their μ-DS. Real radar data from a Thales Man-portable Surveillance and Target Acquisition Radar (MSTAR) radar was used to test the DTW classifier. The dataset contained examples of the three classes listed above travelling at various headings relative to the radar and it was augmented with two synthesized "unknown" targets. The data was classified by DTW in both the time and frequency domain.

When performing classification in the frequency domain, suitable features for classification may be extracted from a radar target's μ-DS by a combination of ground clutter removal, Discrete Fourier Transform (DFT), frequency normalization and Principal Component Analysis (PCA). This processing was described in detail in Chapter 5 and it was successfully applied for the naïve Bayesian classifier in Chapter 7.

A naïve Bayesian classifier, combined with suitable feature extraction and an "unknown" threshold applied to its raw scores, can discriminate between personnel, wheeled vehicles, tracked vehicles and "unknown" inputs based on the μ-DS. In Chapter 7 the results of classification of the Thales MSTAR data by the naïve Bayesian classifier are presented. The results indicate that the combination of the classifier with the feature extraction process resulted in a particularly high performance radar ATR solution.

The performance of the naïve Bayesian system when classifying the three ground targets of the Thales MSTAR dataset is close to the maximum predicted by the method proposed in Chapter 3 under certain circumstances. It is necessary that sufficient variance remains in the data following feature extraction and that sufficient target aspects are included in the reference dataset. The dependence on number of aspects in the reference dataset prompted the investigation of multistatic micro-Doppler and multiperspective classification in the Chapter 8.

The theory of the micro-Doppler effect can be readily extended to include multistatic radar sys-

tems. By considering a multistatic radar as comprised of a number of bistatic and monostatic channels it is possible to describe the μ-DS present in each channel mathematically for simple targets. The proposed methodology was used to develop a simplified model of a personnel target being illuminated by a multistatic radar.

The use of the multistatic personnel μ-DS model demonstrated that the μ-DS data recorded by reciprocal channels of a multistatic radar is highly correlated. This correlation can be considered as leading to information redundancy since highly correlated data does not provide further useful information about a target's class. From this it was speculated that by removing redundant channels it would be possible to reduce the workload of the data fusion process required for multiperspective classification.

It is worth noting that not all of the novel contributions listed in section 1.2 have resulted in a definite conclusions. Of particular significance is the lack of conclusions regarding the use of correlation distance to select entries for the reference dataset and the use of multiperspective classification for μ-DS based radar ATR. It was felt that while both of these areas are novel contributions to the field their investigation was insufficient for firm conclusions. In the case of the use correlation distance for dataset creation this was a result of the requirements of the k-Nearest Neighbour (k-NN) classifier. As discussed in section 6.3 it was necessary to overpopulate the reference dataset so that k-NN could operate with it. Following from this the pure correlation distance method was not investigated thoroughly since it was desired that the datasets should be kept consistent between the different tests and classifiers to allow fair comparison of the results. With regard to the importance of multiperspective classification, it was apparent in the discussion of sections 8.6 and 8.7 that the vehicle data gathered using NetRAD was unsuitable for rigorous testing. These two areas then suggest that further investigation of the subject matter in this thesis is possible.

9.2 Future Work

Like all research, the investigation reported in this document suggests further work that could be performed in the same field. This future work can be divided into three broad areas.

9.2.1 Further Methodology Development

As part of the "black box" classifier a "don't declare" threshold was proposed, but it was not considered further. The threshold was to be responsible for identifying situations in which the scores output by the classifier were too close together to make a reliable declaration for one target class or another i.e. to check for draws. It was not investigated during the current research as it placed demand for even more data on the investigation than the extra evaluation metrics and the "unknown" threshold required. It was felt there was insufficient data available to support all the new initiatives so investigation of the effects of a "don't declare" threshold were left for the future. It therefore forms an obvious first choice for future

developments.

The method for estimating the classifier performance proposed was based on the Bhattacharyya bound on the error and as such makes several simplifications and assumptions. The use of numerical differentiation methods would allow the more accurate Chernoff bound to be used, however, it would still be necessary to perform pairwise comparisons and to assume normal distribution of the data. The greatest step that could be taken in this area would be to investigate the use of the separability index (Oh et al., 1999; Mishra and Mulgrew, 2008) that can be calculated for multiple classes with resorting to binary decisions. Unfortunately, the use of such an index will greatly increase the complexity of estimating the performance since it will require the evaluation of the probability density functions for target classes in feature space.

9.2.2 Monoperspective μ-DS Radar ATR Investigation

Further testing of the correlation distance method for reference dataset creation could be performed to contrast it with the common random selection method. The necessity of having datasets that could be used with all the classifiers investigated, so the results could be compared, limited the research into using correlation distance to select data for the reference dataset. The sparse dataset testing of Chapter 6 demonstrates that the approach works successfully but does not compare it with other approaches. Conducting tests to compare the correlation distance method with the random selection method would allow clear analysis about how each approach effected dataset size and radar ATR system performance. Following the confirmation of the dataset selection process more detailed analysis of the classifiers could be performed.

The use of PCA to reduce the dimensionality of the feature vectors could be applied to the data presented to the DTW and k-NN classifiers. Such additional pre-processing would potentially remove the problems identified in section 6.4 where the classifiers were confused by the noise regions of the feature vectors. Such improvements to feature extraction would be relatively straightforward to implement: the decision to not investigate this idea as part of this research was based purely on time constraints. Further to these simple developments to the monoperspective classification work, the naïve Bayesian classifier could be enhanced by changing its implementation.

During the development of the naïve Bayesian classifier it was assumed that the elements of the feature vectors were normally distributed. This simplification reduces the accuracy of the classifier but makes implementation easier. With the merit of this classifier clearly identified by this research it is apparent that better estimation of the probability distributions would be worthwhile. Potential numerical estimation methods that could be employed are parzen windows (Duda et al., 2005) and kernel density estimators (Botev, 2006) both of which allow the curve of the probability distribution function to be found without knowing the underlying function. It would be expected that with such approaches in place

the distinction between known and unknown classes would be greater and that the results would better match those estimated by either Bhattacharyya bound or the separability index. While it may appear from this research that the classifier is already close to its theoretical best, as more classes as added the problem of discriminating between them becomes harder and is likely that suggested improvement would be required.

It was mentioned in section 6.4 that the Thales MSTAR ATR system uses the classifications from successive scans to improve confidence in the declared class of a target (Stove and Sykes, 2003). None of the work presented in this thesis exploits the scanning nature of typical radar functions. Development of similar algorithms may improve the performance of all the classifiers considered. As such, investigation of this functionality would be a particularly significant piece of future work.

9.2.3 Multiperspective μ-DS Radar ATR Investigation

Future work for the multiperspective radar ATR research can be broken into three areas: further simulation work; further multistatic μ-DS trials and analysis; and further classification research.

The multistatic μ-DS simulation research could be extended in two ways. Firstly, more advanced models for human locomotion could be used to produce increasingly realistic simulated μ-DS in each channel of the multistatic radar. Essentially the work of authors such as Ram and van Dorp (van Dorp and Groen, 2003; Ram and Ling, 2008) would be extended into the multistatic domain. Secondly, other target classes could be simulated. While it may be difficult to develop the mathematical models required for simulations of a complex target, such as a tracked vehicle, Ram's approach provides an alternative. Instead of using mathematical formulae, her simulations are based on target models developed in computer animation software. Using this technique it would be possible to draw, then animate and finally derive the μ-DS of a complex target without needing the mathematical description of each subcomponent's motion. The development of such simulations would be of great support to further trials work.

The understanding of multistatic μ-DS would be advanced by the collection of more data that could be analysed and used to investigate classification. Without modification of the existing hardware available at University College London (UCL) it would be possible to gather data for targets that were guaranteed to exhibit a strong μ-DS. It was seen in sections 8.6 and 8.7 that both personnel and cyclists had a detectable signatures because the target components with micro-motions were large relative to the overall target Radar Cross Section (RCS). Other targets of this nature, perhaps certain animals, could be identified and used on the trial. A constraint on this approach would be the production of a target dataset with limited interest to conventional radar applications. The dataset could still be used to perform multiperspective radar ATR research although in a proof of concept manner only. Alternatively the radar hardware could be improved to allow collection of the μ-DS for relevant targets.

NetRAD's ability to detect a target's μ-DS could be improved by increasing its transmitted power

and moving to a higher carrier frequency. Such enhancements would allow the collection of a multistatic μ-DS dataset for conventional radar targets. While the transmit power upgrade is already being performed as part of a multistatic sea clutter investigation, the carrier frequency change would prove difficult. To move to a frequency in the Ku-band, say, would require substantial hardware re-design and a broadcast license from OfCom, the United Kingdom's spectrum licensing authority. However, assuming suitable hardware upgrades were made and a license obtained it would be possible to gather a comprehensive multistatic μ-DS dataset with relative ease. Such a dataset would allow detailed analysis to be performed and classification investigations to be undertaken.

The first step to performing further research into multistatic μ-DS based radar ATR is to gather a high quality test dataset. Either of the two approaches suggested above, using NetRAD and targets guaranteed to exhibit detectable signatures or developing improved hardware to detect signatures on conventional targets, would suffice. It may be possible to simulate test data but if a large number of target classes are required it may be more time consuming to simulate the data than to gather it at a trial. Once the dataset has been obtained there are three critical areas for further investigation.

Firstly, the way in which features can be extracted from the multistatic data will require investigation. The method proposed in section 5.4 could be applied to the data from each multistatic channel in turn or across all the channels simultaneously. The advantages and disadvantages of each possibility are unclear and merit further investigation. Furthermore, there are known differences between the channels, such as the size of the Doppler shift and the bandwidth of the μ-DS, whether these should be accounted for during feature extraction should also be researched.

Secondly, the data fusion process, where the different channels are fused into a single feature vector, can be improved above that presented in section 8.7. In this work the DTW algorithm's inbuilt ability to handle multidimensional data was used to provide data fusion, however many classifiers do not have this ability so a separate stage must be developed. It was identified in sections 8.4 and 8.6 that there was redundancy in the channels of the multistatic radar data. Investigating how this can be exploited during data fusion would be a key part of any future work in this area.

Thirdly, as was identified in section 8.7, the DTW classifier requires that the data channels be presented in a fixed order and this may limit its ability to process multiperspective data. Other classifiers have been proposed that do not have this requirement (Vespe et al., 2007) and the investigation of these will be essential to understand the possibilities offered by multistatic μ-DS data for radar ATR. Of particular interest is the implementation of the naïve Bayesian classifier Vespe proposes since the naïve Bayesian classifier was observed as being particularly effective at classifying monostatic μ-DS data.

It is important to realize that the three parts of the suggested further research into multiperspective μ-DS classification are strongly connected. The classifier being used may effect the feature extraction process required and this could, in turn, impact on the data fusion process. Investigating how the three

areas effect each other would then form a forth part of this future multistatic μ-DS classification research.

9.2.4 Advanced Analysis Of The μ-DS

In section 4.3 the manner in which μ-DS data would be analysed in the thesis was constrained to the use of the DFT and the Short Time Fourier Transform (STFT). The decision to use the STFT as the sole time-frequency transform was somewhat limiting since the reduced durations of the DFTs utilized in the transform resulted in low frequency resolution. While the conventional high-resolution time-frequency transforms are limited for μ-DS work by their cross-terms (Kootsookos et al., 1992) there are some advanced methods that retain a high enough dynamic range to be valuable (Marple, 2003). Utilizing these in the study of the data collected using the NetRAD system may allow the observation of signature features close to the Main Doppler Line (MDL) that were obscured by the broadening of the line in the STFT used in the presented investigation. As such, including them in the future work would form a valuable extension to this study.

Appendix A

Communication With Prof. David Hand

The following e-mail exchange was made with Prof. David Hand of Imperial College London to support use of the naïve Bayesian classifier in Chapter 7.

Subject: Question following from your paper "Idiot's Bayes–Not So Stupid After All?"

Date: 08/12/2007

Dear Prof. Hand,

I recently read your 2001 paper *Idiot's Bayes—Not So Stupid After All?* in relation to some research I am conducting into radar target classification: I found the paper both useful and interesting but it did leave me with a lingering question and I wondered what your view on it would be. If part of the pre-processing of the data includes principal component analysis does the naive assumption that the elements of the feature vector are statically independent become a certainty? Based on my understanding of the uses of PCA once the original data has been projected onto the main principal components the elements of the new feature vectors should be orthogonal and therefore independent. Under such conditions an assumption of naivety is no longer naive, but rather sensible. I have struggled to find a satisfactory answer to this question, and it is while looking for one that I came across your paper.

Just to give you a little background if you are interested: my research is for a PhD and I am attempting to recognizing the class of radar targets based around Doppler characteristics that arise in the backscattered electromagnetic signal. Like many of the researchers you reference in your paper I have experienced great success with a naive Bayes classifier despite the naivety of the underlying assumptions. I began using PCA to help reduce the dimensionality of my data, since it typically ranges from 64 to 512 elements per sample, and because it was apparent from underlying theory that there would be dependencies. While thinking about why the naive Bayes classifier had been so successful I realized that the PCA was acting to make the elements of the feature vectors independent. Surprisingly however I have not been able to find any published material on the combined use of PCA and naive Bayes classification.

I look forward to your reply and would be happy to answer any questions you may have about my research.

Your sincerely,

Graeme

Subject: Re: Question following from your paper "Idiot's Bayes–Not So Stupid After All?"

Date: 08/12/2007

Dear Graeme,

You are right. The principal components are independent, so if one transforms to those and uses them as the variables in a naive Bayes classifier, then the assumption is reasonable. Of course, there is some uncertainty due to the fact that the PCs will be computed from a sample, rather than from the true distributions, but this will generally be unimportant.

However, I think there is a deeper issue. The naive Bayes method assumes independence of the variables within each class. If the PCs were the same in the different classes, then the transformations to PCs would be the same, and what you say would follow. However, if the PCs were different, then we would be stuck with a model which effectively used different variables for the different classes.

There is work on 'common principal components' which forces the extracted variables to be the same in the different classes. Presumably, though, this will normally result in "PCs" which do have some residual correlation. See, for example, Bernhard Flury (1988) Common Principal Components and Related Multivariate Models. Wiley. It is also discussed in I. T. Jolliffe (2002) Principal Component Analysis. Springer.

Regards

David Hand

Appendix B

Communication With Prof. Eamonn Keogh

The following e-mail exchange was made with Prof. Eamonn Keogh of the University Of California with regard to the DTW classifier used in Chapters 6 and 7. The paper refered to in the first e-mail is (Smith et al., 2006c).

Subject: Dynamic time warping paper that might interest you

Date: 22/01/2007

Dear Eamonn,

Just over a year ago I e-mailed you asking for a reference to one of your DTW papers that I'd found through google. In your reply you requested that if I ever wrote a paper you'd like to see a copy. Now that I'm well underway on my PhD I've actually written three papers for conferences and thought you might like to see the most recent one. (I'm only sending the one as it's a kind of summary of the other two that I put together for an IET seminar day.) The idea is to use DTW to recognize radar targets based on the way the target modulates the radar signal during reflection. In our original e-mail exchange you mentioned that you'd done some investigation of how bats identify targets. My group here at UCL is actually looking at bats too and we suspect that they're using the same type of phenomenon as the paper I've attached uses for the basis of their classification.

I hope you find it interesting.

Yours sincerely,

Graeme

Subject: Re: Dynamic time warping paper that might interest you

Date: 22/01/2007

Hi Graeme, this is a very nice paper. Some good news! With a tiny bit more work, you can make DTW

both much faster and more accurate!

To make DTW more accurate find the best warping window width. See section 4 of [a].

To make DTW faster, use LB_Keogh lower bounding, and use "early abandoning" ([b] definition 1). I guarantee you that it will be at least two orders of magnitude faster. In Matlab, this would mean adding a total 3 lines of code [c].

In particular, you should be able to get DTW to within about only twice as slow as Euclidean Distance as the datasets get bigger.

Again, nice paper.

Eamonn

[a] *Ratanamahatana, C. A. and Keogh. E. (2004). Everything you know about Dynamic Time Warping is Wrong. **Third Workshop on Mining Temporal and Sequential Data, in conjunction with the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), August 22-25, 2004 - Seattle, WA*

[b] Eamonn Keogh, Li Wei, Xiaopeng Xi, Sang-Hee Lee and Michail Vlachos (2006) *LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures.* VLDB 2006.

[c] `http://www.cs.ucr.edu/~eamonn/LB_Keogh.htm`

Appendix C

Resulting Publications

This research has resulted in the following publications:

Naive Bayesian Radar Micro-Doppler Recognition,
Graeme E. Smith, Karl Woodbridge and Chris J. Baker,
The International Radar Conference 2008, Adelaide. 2008

Radar Classification Evaluation,
Graeme E. Smith, Michele Vepse, Karl Woodbridge and Chris J. Baker,
Proceedings Of The 2008 IEEE Radar Conference, Rome, pp.1585-1590. 2008

Multistatic Micro-Doppler Signature Of Personnel,
Graeme E. Smith, Karl Woodbridge and Chris J. Baker,
Proceedings Of The 2008 IEEE Radar Conference, Rome, pp.1961-1966. 2008

Multiperspective Micro-Doppler Signature Classification,
Graeme E. Smith, Karl Woodbridge and Chris J. Baker,
Proceedings of the IET International Conference on Radar Systems 2007, Edinburgh. 2007

Template Based Micro-Doppler Signature Classification,
Graeme E. Smith, Karl Woodbridge and Chris J. Baker,
Proceedings Of The 3rd European Radar Conference, Manchester, pp.158–161. 2006

Micro-Doppler Signature Classification,
Graeme E. Smith, Karl Woodbridge and Chris J. Baker,
Proceedings Of The 2006 CIE International Radar Conference, Shanghai, pp.1436–1439. 2006

Template Based Micro-Doppler Signature Classification (Expanded Version),
Graeme E. Smith, Karl Woodbridge and Chris J. Baker,
IET High Range Resolution Imaging And Target Classification Seminar Day, November 2006, London, pp.129–134. 2006

Bibliography

Bell, M. and Grubbs, R. (1993), JEM modeling and measurement for radar target identification, *IEEE T. Aero. Elec. Sys.* **29**(1), 73–87.

Bilik, I., Tabrikaian, J. and Cohen, A. (2006), GMM-based target classification for ground surveillance Doppler radar, *IEEE Transactions on Aerospace and Electronic Systems* **42**(1), 267–278.

Bishop, C. M. (2005), *Pattern RecognitionAndMachineLearning*, draft 1.22 edition, Springer Science And Business Media, Inc.

Botev, Z. I. (2006), A Novel Nonparametric Density Estimator, Technical Report, The University of Queensland, Queensland, Austalia.

Boulic, Ronan; Thalmann, N. M. T. D. (1990), A global human walking model with real-time kinematic personification, *The Visual Computer* **6**(6), 344–358.

Britannica (2007), *Encyclodpædia Britannica*, Encyclodpædia Britannica Online, Chapter Radar, pp.http://search.eb.com/eb/article–28736.

Brookner, E. (2006), Phased Arrays and Radars—Past, Present and Future, *Microwave Journal* **49**(1), http://www.mwjournal.com/article.asp?HH_ID=AR_29.

Chen, V. (2000), Analysis of radar micro-Doppler with time-frequency transform, *in Statistical Signal and Array Processing, 2000.Proceedings of the Tenth IEEE Workshop on*, pp.463–466.

Chen, V. C. (2008), Detection And Analysis Of Human Motion By Radar, *in Proceedings Of The 2008 IEEE Radar Conference*, Rome, pp.1957–1960.

Chen, V. C., Fayin, L. and Shen-Shyang, H. (2006), Micro-doppler effect in radar: phenomenon, model, and simulation study, *IEEE Transactions on Aerospace and Electronic Systems* **42**(1), 2–21.

Chen, V. C. and Ling, H. (2002), *Time-Frequency Transforms For Radar Imaging And Signal Analysis*, 1 edition, Artech House.

Chen, V., Li, F., Ho, S. and Wechsler, H. (2003), Analysis of micro-Doppler signatures, *Radar, Sonar and Navigation, IEE Proceedings* - **150**(4), 271–276.

Cherniakov, M. (2007), *Bistatic Radar Principles And Practice*, John Wiley and Sons, Ltd.

Chernyak, V. S. (1998), *Fundamentals Of Multisite Radar Systems (Multistatic Radars And Multiradar Systems)*, Gordon And Breach Scientific Publishers.

Cohen, M. (2006), Radar and Multi-Sensor Target Recognition, *in 2006 IEEE Radar Conference (Tutorials)*, Verona, NY.

Cooley, J. W. and Tukey, J. W. (1965), An Algorithm for the Machine Calculation of Complex Fourier Series, *Mathematics of Computation* **19**(90), 297–301.

Cuomo, S., Pellegrini, P. and Piazza, E. (1994), Model validation for 'jet engine modulation' phenomenon, *Electronics Letters* **30**(24), 2073–2074.

Deng, H. (2004), Polyphase code design for Orthogonal Netted Radar systems, *Signal Processing, IEEE Transactions on* **52**(11), 3126–3135.

Derham, T., Doughty, S., Woodbridge, K. and Baker, C. (2007), Design and evaluation of a low-cost multistatic netted radar system, *Radar, Sonar & Navigation, IET* **1**(5), 362–368.

Derham, T., Doughty, S., Woodbridge, K. and Baker, C. J. (2006), Realisation and Evaluation of a Low Cost Netted Radar System, *in Proceedings Of The CIE 06 International Conference on Radar*, pp.1–4.

Doughty, S. (2008), Development And Performance Evaluation Of A Multistatic Radar System, PhD thesis, University College London, London.

Doughty, S. R., Woodbridge, K. and Baker, C. J. (2007), Improving Resolution Using Multistatic Radar, *in Proceedings of the IET International Conference on Radar Systems2007*, Edinburgh.

Duda, R. O., Hart, P. E. and Stork, D. G. (2005), *Pattern Classification*, 2 edition, John Wiley & Sons Inc.

Egan, J. P. (1975), *Signal Detection Theory And ROC Analysis*, Academic Press, Inc (London) Ltd.

Fawcett, T. (2003), ROC Graphs: Notes and Practical Considerations for Data Mining Researchers, Technical Report HPL-2003-4, HP Laboratories, Intelligent Enterprise Technologies Laboratory HP Laboratories Palo Alto.

Ferri, M., Serrano, F., Dainelli, V. and Galati, G. (2003), Millimeter wave technologies for Airport Surface Movement Control Applications, *in Proceedings of 3rd ESA Workshop on Millimetre Wave*

Technology and Applications: Circuits, Systems, and Measurement Techniques (ESA-WPP-212), ESA, Oerlikon Contraves SpA, Rome, Italy;, pp.589–594.

Gamba, P., Houshmand, B. and Saccani, M. (2000), Detection and extraction of buildings from interferometric SAR data, *Geoscience and Remote Sensing, IEEE Transactions on* **38**(1), 611–617.

Gardner, R. E. (1961), Report 5656, Technical Report, Naval Research Laboratory, Washington, DC.

Gauss, C. F. (1866), Nachlass: Theoria interpolationis methodo nova tractata, *in Carl Friedrich Gauss, Werke, Band 3*, Gottingen: Koniglichen Gesellschaft der Wissenschaften, pp.265–303.

Ghaleb, A., Vignaud, L. and Nicolas, J.-M. (2008), Micro-Doppler analysis of pedestrians in ISAR imaging, *in Proceedings Of The 2008 IEEE Radar Conference*, Rome, pp.943–947.

Hai, D. (2004), Effective CLEAN algorithms for performance-enhanced detection of binary coding radar signals, *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]* **52**(1), 72–78.

Hand, D. J. and Yu, K. (2001), Idiot's Bayes: Not So Stupid after All?, *International Statistical Review / Revue Internationale de Statistique* **69**(3), 385–398.

Hecht, E. (1987), *Optics*, 2 edition, Addison-Wesley Publishing Company.

Högbom, J. (1974), Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines, *Astronomy and Astrophysics Supplement,* **Vol**, 417–426.

Hrynkiewicz, E. (1998), Rectangular function Π(x) and some of its applications for digital circuits design, *in Electronic Instrument Engineering Proceedings, 1998. APEIE-98. Volume 1. 4th International Conference on Actual Problems of*, pp.386–390.

Hynes, R. and Gardner, R. E. (1967), Doppler Spectra Of S Band And X Band Signals, *IEEE Transactions on Aerospace and Electronic Systems* **AES-3**(6), 356–365.

Jahangir, M., Blacknell, D., Moate, C. and Hill, R. (2007), Extracting information from shadows in SAR imagery, *in International Conference On Machine Vision*, pp.107–112.

Jahangir, M., Ponting, K. and O'Loghlen, J. (2003a), Correction to 'Robust Doppler classification technique based on hidden Markov models', *Radar, Sonar and Navigation, IEE Proceedings* - **150**(5), 387.

Jahangir, M., Ponting, K. and O'Loghlen, J. (2003b), Robust Doppler classification technique based on hidden Markov models, *Radar, Sonar and Navigation, IEE Proceedings* - **150**(1), 33–36.

Jiajin, L. (2005), Pattern recognition based on time-frequency distributions of radar micro-Doppler dynamics, *in Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks.SNPD/SAWN 2005.Sixth International Conference on*, pp.14–18.

Jiajin, L. and Chao, L. (2005), Target classification based on micro-Doppler signatures, *in 2005 IEEE International Radar Conference Record (IEEE Cat. No.05CH37628)*, IEEE, Dept. of Comput. & Inf. Sci., Towson Univ., MD, USA, pp.179–183.

Jofre, L. (1993), Research and new developments in radar cross section prediction and measurement in Spain, *in Proceedings of the 23rd European Microwave Conference*, *European Microwave Conference, 1993. 23rd* , pp.42–46.

Johnsen, T., K.E., O. and Gundersen, R. (2003), Hovering helicopter measured by bi-/multistatic CW radar, *in Proceedings Of The 2008 IEEE Radar Conference*, pp.165–170.

Kailath, T. (1967), The Divergence and Bhattacharyya Distance Measures in Signal Selection, *Communications, IEEE Transactions on* **15**(1), 52–60.

Keogh, E. J. and Pazzani, M. J. (2001a), Derivative Dynamic Time Warping, *in First SIAM International Conference on Data Mining (SDM'2001)*, Chicago.

Keogh, E. J. and Pazzani, M. J. (2001b), Dynamic Time Warping with Higher Order Features, *in First SIAM International Conference on Data Mining (SDM'2001)*, Chicago.

Keogh, E. J. and Ratanamahatana, C. A. (2002), Exact Indexing Of Dynamic Time Warping, *in 28th International Conference on Very Large Data Bases*, Hong Kong, pp.406–417.

Kim, K.-T., Seo, D.-K. and Kim, H.-T. (2005), Efficient classification of ISAR images, *Antennas and Propagation, IEEE Transactions on* **53**(5), 1611–1621.

Kingsley, S. and Quegan, S. (1992), *Understanding Radar System*, McGraw-Hill International (UK) Ltd.

Knott, E. F., Shaeffer, J. F. and Tuley, M. T. (2004), *Radar Cross Section*, Vol. 2, SciTech Publishing Inc.

Kootsookos, P., Lovell, B. and Boashash, B. (1992), A unified approach to the STFT, TFDs, and instantaneous frequency, *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]* **40**(8), 1971–1982.

Levanon, N. and Mozenson, E. (2004), *Radar Signals*, John Wiley.

Liao, X., Bao, Z. and Xing, M. (2000), On the aspect sensitivity of high resolution range profiles and its reduction methods, *in The Record of the IEEE 2000 International Radar Conference*, pp.310–315.

Liao, X., Runkle, P., Jiao, Y. and Carin, L. (2001), Identification of ground targets from sequential HRR radar signatures, *in Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on* **5**, pp.2897–2900 vol.5.

Marple, S.L., J. (2001), Large dynamic range time-frequency signal analysis with application to helicopter Doppler radar data, *in Signal Processing and its Applications, Sixth International, Symposium on. 2001*, Vol. 1, pp.260–263.

Marple, S.L., J. (2003), Sharpening and bandwidth extrapolation techniques for radar micro-Doppler feature extraction, *in Radar Conference, 2003.Proceedings of the International*, pp.166–170.

Martin, J. and Mulgrew, B. (1990), Analysis of the theoretical radar return signal form aircraft propeller blades, *in Radar Conference, 1990., Record of the IEEE 1990 International*, pp.569–572.

Martin, J. and Mulgrew, B. (1992), Analysis of the effects of blade pitch on the radar return signal from rotating aircraft blades, *in Radar 92.International Conference*, pp.446–449.

Martinez Madrid, J., Casar Corredera, J. and de Miguel Vela, G. (1992), A neural network approach to Doppler-based target classification, *in Radar 92.International Conference*, pp.450–453.

Mishra, A. K. and Mulgrew, B. (2008), Multipolar SAR ATR: Experiments with the GTRI Dataset, *in Proceedings Of The 2008 IEEE Radar Conference*.

Misiurewicz, J., Kulpa, K. and Czekala, Z. (1998), Analysis of radar echo from a helicopter rotor hub, *in Microwaves and Radar, 1998.MIKON '98., 12th International Conference on*, Vol. 3, pp.866–870.

Mulgrew, B., Grant, P. and Thompson, J. (2003), *Digital Signal Processing Concepts And Applications*, second edition, Palgrave Macmillan.

Myers, C., Rabiner, L. and Rosenberg, A. (1980), Performance tradeoffs in dynamic time warping algorithms for isolated word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**(6), 623–635.

Oh, I.-S., Lee, J.-S. and Suen, C. (1999), Analysis of class separation and combination of class-dependent features for handwriting recognition, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **21**(10), 1089–1094.

Pellegrini, P., Cuomo, S. and Piazza, E. (1995), ATC primary radar signal analysis for target characterisation. A model validation, *in Radar Conference, 1995., Record of the IEEE 1995 International*, pp.516–520.

Perkins, L., Smith, H. and Mooney, D. (1984), The Development of Airborne Pulse Doppler Radar, *Aerospace and Electronic Systems, IEEE Transactions on* **AES-20**(3), 292–303.

Piazza, E. (1999), Radar signals analysis and modellization in the presence of JEM application to civilian ATC radars, *Aerospace and Electronic Systems Magazine, IEEE* **14**(1), 35–40.

Proakis, J. G. and Manolakis, D. G. (1996), *Digital Signal Processing—Principles, Algorithms, And Applications*, 3rd edition, Prentice-Hall, Inc, New Jersey, USA.

Ram, S. S. and Ling, H. (2008), Simulation of Human Doppler Spectrograms, *in Proceedings Of The 2008 IEEE Radar Conference*, Rome, pp.535–540.

Rihaczedk, A. W. and Hershkowitz, S. J. (2000), *Theory And Practice Of Radar Target Identification*, 1 edition, Artech House.

Robinson, M., zimi Sadjadi, M. and Salazar, J. (2005), Multi-aspect target discrimination using hidden Markov models and neural networks, *Neural Networks, IEEE Transactions on* **16**(2), 447–459.

Runkle, P., Carin, L. and Lam, N. (2000), Multi-aspect target classification using hidden Markov models for data fusion, *in Geoscience and Remote Sensing Symposium, 2000.Proceedings.IGARSS 2000.IEEE 2000 International*, Vol. 5, pp.2123–2125.

Sakoe, H. and Chiba, S. (1978), Dynamic programming algorithm optimization for spoken word recognition, *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on* **26**(1), 43–49.

Sammartino, P. F., Baker, C. J. and Rangaswamy, M. (2008), Moving target localization with multistatic radar systems, *in Proceedings Of The 2008 IEEE Radar Conference*, Rome, pp.512–517.

Sankoff, D. and Kruskal, J. (1999), *Time Warps, String Edits, And Macromolecules—The Theory And Practice Of Sequence Comparison*, The David Hume Series, CSLI Publications.

Scarth, R. N. (1999), *Echoes From The Sky: A Story of Acoustic Defence*, Hythe Civic Society.

Schneider, H. (1987), On the maximum entropy method for Doppler spectral analysis of radar echoes from rotating objects, *in International Conference Radar 87 (Conf. Publ. No.281)*, IEE, London, UK, pp.279–282.

Schumacher, R. and Schiller, J. (2005), Non-cooperative target identification of battlefield targets - classification results based on SAR images, *in Proceedings Of The IEEE International Radar Conference*, pp.167–172.

Showman, G., Richards, M. and Sangston, K. (1998), Comparison of two algorithms for correcting zero-Doppler clutter in turntable ISAR imagery, *in Signals, Systems & Computers, 1998.Conference Record of the Thirty-Second Asilomar Conference on*, Vol. 1, pp.411–415.

Silverstein, P., Sands, O. and Garber, F. (1991), Radar target classification and interpretation by means of structural descriptions of backscatter signals, *Aerospace and Electronic Systems Magazine, IEEE* **6**(5), 3–7.

Skolnik, M. I. (1980), *Introduction To Radar Systems*, Electrical Engineering Series, 2nd edition, McGraw-Hill.

Skolnik, M. I. (ed.) (1990), *Radar Handbook*, Vol. 2, McGraw-Hill Book Company.

Smith, G. E., Vespe, M., Woodbridge, K. and Baker, C. J. (2008a), Radar Classification Evaluation, *in Proceedings Of The 2008 IEEE Radar Conference*, Rome, pp.1585–1590.

Smith, G. E., Woodbridge, K. and Baker, C. J. (2006a), Micro-Doppler Signature Classification, *in Proceedings Of The 2006 CIE International Conference On Radar*, Vol. 2, Shanghai, pp.1436–1440.

Smith, G. E., Woodbridge, K. and Baker, C. J. (2006b), Template Based Micro-Doppler Signature Classification, *in Proceedings Of The 3rd European Radar Conference*, Manchester, pp.158–161.

Smith, G. E., Woodbridge, K. and Baker, C. J. (2006c), Template Based Micro-Doppler Signature Classification (Extended), *in High Resolution Imaging and Target Classification, 2006. The Institution of Engineering and Technology Seminar on*, London, pp.127–144.

Smith, G. E., Woodbridge, K. and Baker, C. J. (2007), Multiperspective Micro-Doppler Signature Classification, *in Proceedings of the IET International Conference on Radar Systems 2007*, Edinburgh.

Smith, G. E., Woodbridge, K. and Baker, C. J. (2008b), Multistatic Micro-Doppler Signature Of Personnel, *in Proceedings Of The 2008 IEEE Radar Conference*, Rome, pp.1961–1966. 2008.

Smith, G. E., Woodbridge, K. and Baker, C. J. (2008c), Naïve Bayesian Radar Micro-Doppler Recognition, *in Proceedings of the International Radar Conference 2008*, Adelaide.

Stimpson, G. W. (1998), *Introduction To Airborne Radar*, 2 edition, SciTech Publishing Inc.

Stove, A. (2006), A Doppler-Based Target Classifier Using Linear Discriminants and Principal Components, *in High Resolution Imaging and Target Classification, 2006. The Institution of Engineering and Technology Seminar on*, pp.107–125.

Stove, A. G. and Sykes, S. R. (2003), A Doppler-based target classifier using linear discriminants and principle components, *in Proceedings Of The International Radar Conference 2003*, pp.171–176.

Swords, S. S. (1986), *Technical History Of The Beginnings Of RADAR*, IEE History Of Technology Series, Institution Of Electrical Engineers.

Tong, C., Wan, J., Wan, Z. and Wan, G. (1998), Physical mechanism of JEM effect in jet engine, *in Microwave and Millimeter Wave Technology Proceedings, 1998.ICMMT '98.1998 International Conference on*, pp.516–519.

Turhan-Sayan, G. (2005), Real time electromagnetic target classification using a novel feature extraction technique with PCA-based fusion, *Antennas and Propagation, IEEE Transactions on* **53**(2), 766–776.

van Dorp, P. and Groen, F. (2003), Human walking estimation with radar, *Radar, Sonar and Navigation, IEE Proceedings* - **150**(5), 356–365.

Vespe, M. (2007), Multi-Perspective Radar Target Classification, PhD thesis, University College London (University Of London), London.

Vespe, M., Baker, C. J. and d. Griffiths, J. (2007), Radar target classification using multiple perspectives, *Radar, Sonar & Navigation, IET* **1**(4), 300–307.

Vespe, M., Baker, C. J. and Griffiths, H. D. (2006a), Aspect dependent drivers for multi-perspective target classification, *in Radar, 2006 IEEE Conference on*, pp.256–260.

Vespe, M., Baker, C. J. and Griffiths, H. D. (2006b), Outline Structural Representation for Radar Target Classification Based on Non-Radar Templates, *in Radar, 2006. CIE '06. International Conference on*, pp.1–4.

Wenger, J. (2005), Automotive radar - status and perspectives, *in Compound Semiconductor Integrated Circuit Symposium, 2005. CSIC '05. IEEE*, pp.21–24.

Willis, N. J. (2005), *Bistatic Radar*, online edition, SciTech Publishing.

Zhang, H. (2004), The Optimality of Naïve Bayes, *in Proceedings of the 17th International FLAIRS conference*, Miami Beach, Florida.

Zhu, F., Zhang, X.-D., Hu, Y.-F. and Xie, D. (2007), Nonstationary Hidden Markov Models for Multi-aspect Discriminative Feature Extraction From Radar Targets, *IEEE T. Signal. Proces.* **55**(5), 2203–2214.

Zyweck, A. and Bogner, R. (1995), Coherent averaging of range profiles, *in Record of the IEEE 1995 International Radar Conference*, *Radar Conference, 1995., Record of the IEEE 1995 International* , pp.456–461.