High Sensitivity Optical Detection Using Temporal Coherence Interferogram Phase Changes

by

Ricardo Cesar Coutinho

Thesis Presented for Examination
for the Degree of Ph.D. at the
University of London

Department of Electronic and Electrical Engineering

University College London

January 2003

ACKNOWLEDGEMENTS

I am deeply thankful to Prof. Herbert A. French (OBE) not only for giving me the opportunity to explore one of his patents in this research, but also and mainly for teaching my first lessons in optics and in coherence theory. He was the best teacher I could ever have for these subjects because, as happened to him in the nineteen sixties, I have also moved from a radar and signal processing background to the optics field.

To my supervisor Prof. Hugh D. Griffiths, for his firm supervision, being always supportive since I left my country in 1998, and for believing in my capacity to succeed.

To my subsidiary supervisor David R. Selviah, for stretching me to the limit of my work capacity, for continuously seeking perfection, and for his wise help in the improvement of my writing in English.

To Duleep Wickramasinghe for fruitful tuition and discussions, and for providing most of the equipment used in this thesis, with permission from DERA Portsdown West.

To my colleagues in the two research groups I have been involved with, especially Lawrence Commander from the Optical Systems and Devices Group, and Richard Bullock from the Microwaves, Radar and Optics Group, for telling me the "how to"s.

To the Brazilian Navy, for sponsoring my studies in the UK for 28 months.

To my wife and children, for sharing their husband and father with such a time-consuming rival without giving me up.

To God, for allowing all of this to happen.

Finally, I wish to dedicate this thesis to my mother, who has given me support and encouragement to strive the long road towards a PhD, but unfortunately did not live to share the joy of the achievement with me.

ABSTRACT

Most current detection and imaging systems rely on intensity differences between the object of interest and the background. With the recent advances in detector and computing technologies, sensitivities are reaching the background quantum fluctuation level, and performances tend to be limited by background clutter. In order to obtain further contrast enhancement, new dimensionality must be added to the detection process.

This thesis studies a detection technique named Interferogram Phase Step Shift (IPSS), which relies on the coherence contrast between target and background to improve discrimination. The approach employs an interferometer to create a coherence profile, a narrowband optical filter to generate a feature in this profile, and an algorithm to locate this feature. By measuring the shift in this feature in path difference, incremental changes in coherence can be sensed. Unlike Fourier transform spectroscopy, it does not require a Fourier transform, and performs a much narrower path difference scan, generating less data for processing.

Laboratory experiments with different simulated targets and experimental arrangements demonstrate very high sensitivities for coherent sources (e.g. –46 dB signal-to-clutter ratio for a laser), and high sensitivities for partially coherent sources (-30 dB for light from a monochromator). A system based on a CCD camera, with no moving parts, is also shown to yield high sensitivities. The experimental results compare favourably with most competing techniques reviewed. The performance of the proposed detection system was also characterised through the calculation of its Receiver Operating Characteristic (ROC) curves. A laser target with a signal-to-clutter ratio of –26 dB was detected with a probability of detection of 90%, and a probability of false alarm of 2.10^{-4}.

The thesis includes a theoretical model, which predicts the experimental results with good agreement. The theory also calculates the effects of target-to-filter bandwidth ratio and central wavelength offset, which was also confirmed to be in agreement with the experiments.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 5

Chapter 7

LIST OF TABLES

LIST OF ABBREVIATIONS

A/D – Analogue-to-digital

AGC – Automatic gain control

BLIP – Background-limited performance sensor

CCD – Charge coupled device

CCIR – Consultative Committee for International Radio

CLO – Counter low observable

CW – Continuous wave

DBR - Distributed Bragg reflection

DICE – Detection in clutter enhancement

DKLT - Discrete Karhunen-Loève transform

FFT – Fast Fourier transform

FIR – Finite impulse response (filter)

FLIR – Forward-looking infrared

FOV – Field of view

fps – Frame per second

FT – Fourier transform

FTI - Fourier transform image

FTIR - Fourier transform infrared (spectrometer)

FTS - Fourier transform spectroscopy

FWHM – Full width half-maximum

GAP/AN - Glycidyl azide polymer and ammonium nitrate

GFLOPS – Giga-floating point operations per second

GOPS - Giga-operations per second

IC - Imperial College

IFOV – Instantaneous filed of view

IIR - Infinite impulse response (filter)

IPSS – Interferogram phase step shift

IRST – Infrared search and track

LED – Light-emitting diode

LO – Low observable

LOS – Line-of-sight

MBPS – Megabits per second

MCT – Mercury-cadmium telluride

MDSCR – Minimum detectable signal-to-clutter ratio

MLANS – Maximum likelihood adaptive neural system

MRC – Minimum resolvable contrast

MRTD – Minimum resolvable temperature difference

MWIR – Medium wave infrared (band)

NA – Numerical aperture

NECL – Noise-equivalent concentration-path length

NESR – Noise-equivalent spectral radiance

NETD – Noise-equivalent temperature difference

NUC – Non-uniformity correction

OCDR - Optical coherence domain reflectometry

OCT – Optical coherence tomography

OTIM – Optical transform image modulation

PAT – Pointing, acquisition and tracking

PLL – Phase-locked loop

PPM – Pulse position modulation

PR – Power ratio

PSD – Power spectral density

PWM – Pulse width modulation

RCLED – Resonant-cavity light-emitting diode

RIN – Relative intensity noise

RMS – Root-mean square

ROC – Receiver operating characteristic

SCR – Signal-to-clutter ratio

SFTS – Static Fourier transform spectrometer

SIMD - Single instruction, multiple data (computer)

SMF - Spectrally matched filter

SNR – Signal-to-noise ratio

TBD – Track-before-detect

WDM – Wavelength division multiplexing

ZPD – Zero path difference

Chapter 1

Introduction

1.1. Motivation

The most demanding imaging or detection applications of optical techniques are those in which a dim object, weak signal or emission must be located within a scenario of a bright, stronger background. A few examples of this situation are the detection of a tumour in living tissue, of a distant missile in intense daylight, or of a gas in the atmosphere. For the sake of standardisation, in this thesis the object, signal, emission or absorption to be detected is termed *target*, while the radiation coming from the scene is termed *background.* Currently, the majority of the detection or imaging systems employed in these tasks rely on measuring the intensity contrast between the radiation levels received from target and background. Depending on the optical band, different physical mechanisms dominate the radiation received by the detection system. In the visible band, for example, the contrast between light emitted or reflected by the objects in the scene is measured, while in the infrared, the irradiance of these objects dominates. In the former case, reflectivity differences may be important, and in the latter, temperature and emissivity contrast affect target detectability.

In order to face the challenging tasks described, technology evolved in the direction of increasing the sensitivity of the detection systems. The definition of sensitivity used in the thesis is "*the minimum power required to achieve a specified quality of performance in terms of output signal-to-noise ratio or other measure*" [Bod91]. The terms minimum resolvable contrast (MRC), in the visible band, and minimum resolvable temperature difference (MRTD), in the infrared, have been increasingly used, as research efforts push these parameters down to the limit of background quantum fluctuations, characterising a background limited performance sensor (BLIP) [Pol93]. Nevertheless, it is the author's opinion that this quest for the highest sensitivity, using solely the intensity for discrimination, will certainly lead to a dead end. As the sensitivity is increased, clutter structure becomes more visible, and, by using only the intensity for discrimination, performance, instead of improving, may even get worse [Pol00]. Clutter, defined as any target-like features in a scene which are not targets, represents the factor that degrades the performance of imaging systems most severely. Consequently, the weakness of intensity-based systems is the performance in cluttered environments [Moy00].

The dead end just mentioned will not be reached, however, if new dimensions are incorporated to imaging and detection technologies [McC96]. The increase in clutter spatial structure can be used favourably if one records this structure in several different spectral bands, and correlates the acquired images. The additional information contained in the received electromagnetic wave, such as spectrum, polarisation or coherence, is irreversibly destroyed in the intensity detection process. The use of this information before detection allows performance improvement, even in presence of clutter. In fact, techniques employing these domains have been increasingly used in different fields, and some of them are reviewed in Chapter two.

Most detection approaches based on coherence were invented in the 1970s (for example, [Fre72]), and so far have not been widely explored. The term coherence is explained in detail in chapter two, but, in brief, it is the reciprocal of the optical bandwidth. Therefore, techniques relying on coherence can be used to detect targets that present a high degree of coherence (full definition given in chapter three) or that have a narrow optical bandwidth. Man-made targets displaying a high degree of coherence are, for example, lasers and arc lamps. In general, natural objects tend to behave like blackbodies, emitting or reflecting low coherence radiation. However, radiation theory states that rotational and vibrational modes of atoms and molecules produce spectrally selective radiation [Hud69]. The atoms and molecules absorb radiation, vibrate and rotate at frequencies that are a characteristic of the material, and re-radiate electromagnetic waves. These phenomena produce narrow spikes in the emission or absorption spectra of the corresponding materials, which may appear isolated or in bands of spikes. An approach based on coherence, though, may be used to detect man-made or natural targets, provided that they display, in a limited region of their spectra, a narrow spectral spike. In other words, anything that has a narrowband feature in its spectrum is a candidate target for such techniques. A brief list of applications of coherence-based approaches follows, with intent to demonstrate the relevance of the use of coherence.

In the defence field, since the introduction of radar during World War II, there has been a continuous succession of measures, countermeasures and counter-countermeasures aiming to make the two contenders, namely weapons and targets, progressively more capable of detecting and less susceptible to detection by each other. In order to improve their survivability, combat platforms (ships, ground vehicles and aircraft) need to detect their threats (missiles and other platforms) at ranges as long as possible. On the other hand, to ensure lethality, the design of platforms and guided weapons is currently making intensive use of Low Observables (LO) technology [MCT02], also named stealth. In the visible part of the spectrum, camouflage techniques have been widely explored, and the counter-low observables (CLO) measure against it was the use of infrared sensors, which allowed discrimination based on temperature and emissivity differences between target and background.

In the infrared band, the principal sensor installed on defence platforms is the Infrared Search and Track (IRST), which acquires and detects a target based on irradiance differences, mainly through the location of hot spots on its surface. In this band, stealth techniques aim to embed the target in its background, and to avoid the formation of these hot spots by distributing the unavoidable sources of heat across the platform. Further measures to reduce the temperature contrast of aircraft and missiles are the cooling of exhaust gases through mixing with cold air, the use of flameless, low temperature propellants [Rap02] and, for ships, sprinkling seawater on the ship's superstructure [Nee02]. An example of infrared signature reduction is shown in fig. 1.1, taken from [Zis93]. The figure shows the spectral signature of a helicopter at tail aspect, the upper curve being the unsuppressed curve, and the lower curve the suppressed, displaying an intensity level about ten to fifteen times lower. A sensor exploring the target's radiometric signature (total energy irradiated within the sensor's bandwidth) will not be able to discriminate between the suppressed helicopter and the background. This low-contrast situation may happen with unsuppressed targets as well, during twilight, as thermal contrast displays cyclic variations throughout the day [Mou02], and these times of the day represent the points of crossover between positive and negative contrast.

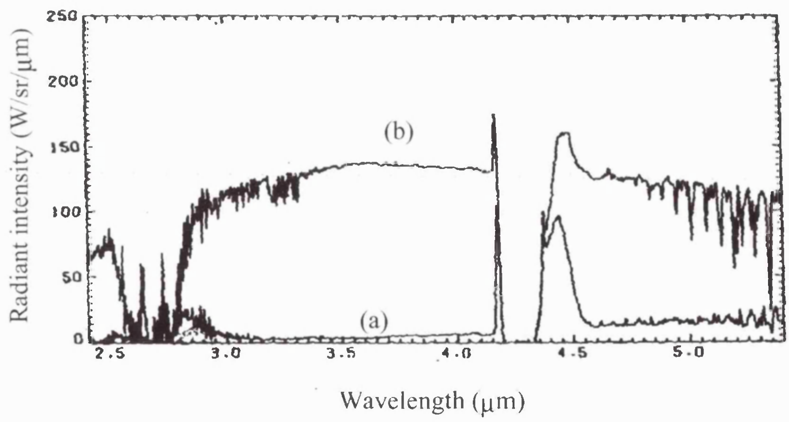


Figure 1.1 – Suppressed (a) and unsuppressed (b) helicopter spectral signature (from [Zis93])

A coherence-based detection system may be used to ensure and even improve the probability of detection, in spite of any target radiometric signature reduction measures. In fact, and for example, fig. 1.1 displays a spectral spike at 4.25 μm in both suppressed and unsuppressed spectra, which is known as the *blue spike*, due to hot CO_2 which is present in plumes from rockets, aircraft and missiles [Cam93]. The phenomenon is created by Doppler broadening of the CO_2 emission due to the high temperature of the plume (from 400 to 2500 K), re-absorbed by colder CO_2 in the atmosphere. As this atmospheric gas is at ambient temperature, this absorption suffers no Doppler spreading, having a narrower band than the emission, hence creating two sidebands which are generated by the hot gas, and not absorbed by the cold gas. By analogy with the visible colours, the band towards smaller wavelengths was named the *blue spike*, while the one with longer wavelength was called the *red spike* or *wing*. It can also be seen in fig. 1.1 that the reduction of the total energy irradiated further increases the spectral contrast between the blue spike and the background continuum. It is reasonable to think that this increased contrast will also happen when the target is on nose aspect, its body blocking the main portion of the radiation, that is, the blackbody-like spectral continuum, unlike the blue spike, which would propagate "*with much less attenuation*" [Zis93]. This difference between target and background attenuation would, then, represent an improvement in signal-to-clutter ratio (SCR).

Real data acquired from the exhaust of a Rolls Royce Spey engine on a test bench placed at 8.6 m from the recording instrument, a Bruker IFS 66 FTIR spectrometer with a spectral resolution of 0.11 cm^{-1}, is shown in fig. 1.2. It shows the blue spike at 4.18 μm, with a full-width half maximum (FWHM) of approximately 30 nm, and an estimated coherence length of 0.6 mm. As will be shown in chapter 8, this spike is perfectly suitable for detection by coherence-based approaches. In fig. 1.2, the red wing can be seen with a band of spikes superimposed, which do not appear in fig. 1.1 due to its lower resolution. The data was kindly provided by Nick Davies, from the Centre for Optical and Environmental Metrology, National Physical Laboratory.

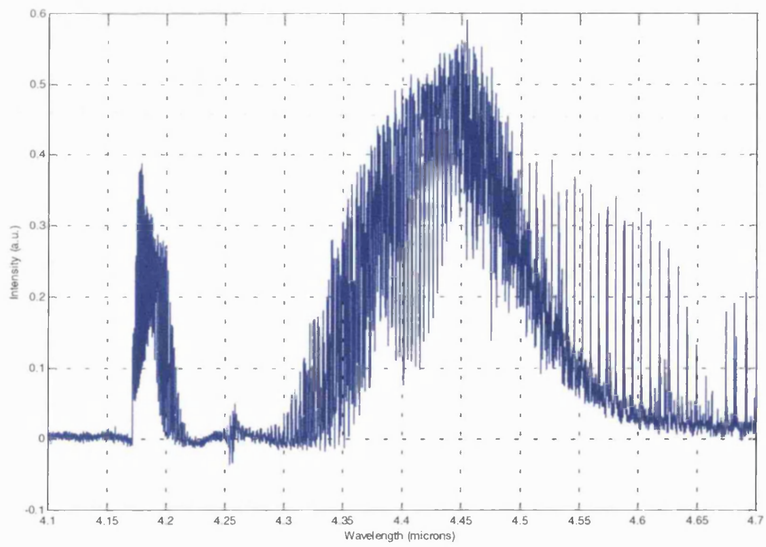


Figure 1.2 - Real data showing the blue spike and red wing effects (from Nick Davies – NPL)

Further narrow spectral features can be found on plumes of jet propelled vehicles. Rapanotti and others [Rap02] have recently published a study on the influence of the chemistry of rocket plumes on their infrared signature. One of the conclusions of the study points to the reduction in the detectability of these targets, using intensity-based infrared sensors, through the use of flameless, low temperature propellants, especially glycidyl azide polymer and ammonium nitrate (GAP/AN). The conclusion is not valid, however, if a coherence sensor is employed. The paper shows that the radiometric-reduced signature contains spectral spikes due to elements not found in the natural atmosphere, once again making rockets using these propellants suitable targets for coherence-based sensors.

It is clear from the previous paragraphs that sensors utilising coherence as a discriminator are suitable for IRST systems, as well as for missile seekers. The author believes, as will be shown in later chapters, that the approach studied in this thesis possesses high sensitivity, making it suitable for target acquisition applications, possibly requiring a handover to conventional sensors as the target closes in, due to its limited dynamic range. Finally, yet one more application for these techniques would be in laser-warning receivers, due to the high degree of coherence of these light sources.

The second application envisaged for coherence-based sensors is the remote sensing of gases, which has already been explored by French [Fre84A] and by Drum [Dru90] in his doctoral thesis, reviewed in chapter two. Contaminant gases like CO, N_2O (used in Drum's measurements), CO_2, H_2CO, $N_2H_3CH_3$ and CH_4 all display selective absorption, forming spectral spikes in the infrared. The range of hydrocarbons found in flammable gases absorbs radiation in bands near 3.3 μm, carbon monoxide at 4.7 μm, nitric oxide at 5.3 μm, and so on. As happens with other spectroscopic techniques employed to detect gases, care must be taken when choosing the spectral region where the sensor is designed to operate. The designer must choose a spectral line, or a small set of lines, away from bands caused by interfering gases, of which the most common is water vapour, whose spectrum spreads widely in the infrared [Hud69].

Another application for sensors exploiting coherence is the detection or imaging of biological specimens. Fig. 1.3 displays the spectrum of deoxyhaemoglobin (Hb) and oxygenated haemoglobin (HbO_2), in the visible band (left) and near infrared (right), obtained from the web-site of the Biomedical Optics Research Group, Department of Medical Physics and Bioengineering, University College London [Elw00]. Hb is found in venous, deoxygenated blood, while HbO_2 is found in arterial blood, after oxygenation by the lungs. The higher absorption coefficient of HbO_2 in blue wavelengths explains the reddish colour of arterial blood, as well as the blue-like colour of venous blood. In the visible band, the author believes, coherence-based approaches could be used to detect and measure the concentration of HbO_2 through its spectral spike at 580 nm. In the near infrared, the same application may be conceived at 760 nm, although the relatively low resolution of the spectra in fig. 1.3 does not allow proper assessment of the feature's coherence length. By measuring the HbO_2 concentration, the blood oxygen level may be determined. Besides real time monitoring of this level, the measurement of blood oxygen level may find application in cancer detection and therapy. Carcinoma cells show higher blood concentrations than normal cells, with stronger absorption in the 500 - 600 nm region [Pet90]. Additionally, the level of oxygen in a tumour affects its resistance to treatment with X-rays, and the treatment has a higher probability of success when this concentration reaches a maximum [OLE98].

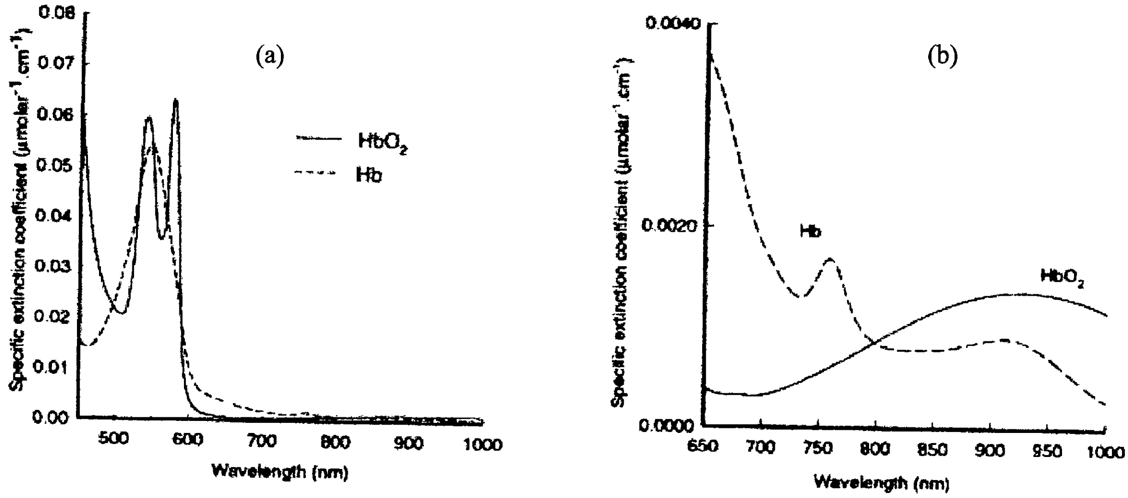


Figure 1.3 - Absorption spectra of Hb and HbO_2 in the visible (a) and near infrared (b)

Forensic science may be another application of coherence-based techniques, in the author's opinion. Fig. 1.4 displays a spectrum of Diazepam measured with an FTIR spectrometer, where a spectral peak free of interfering features is seen at approximately 5.8 µm. The data was kindly offered by the Brazilian Army Research Institute.

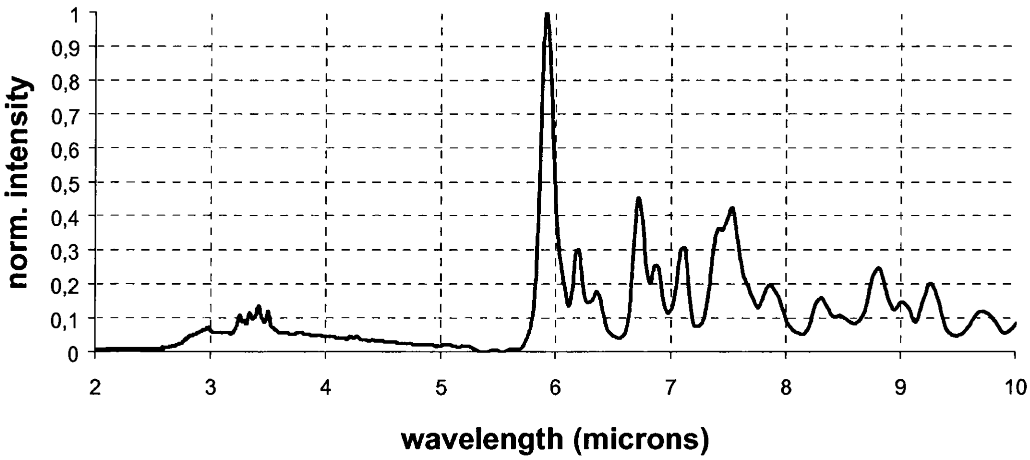


Figure 1.4 - FTIR spectrum of Diazepam (from the Brazilian Army Research Institute)

Yet another interesting application of coherence detection is inter-satellite communications. Constellations of communications or data relay satellites allow the realisation of world-wide networks with increased coverage and very wide bandwidth, for both commercial and military applications. A remarkable example is the real time transmission of imagery from the French SPOT-4 earth observation satellite to the European Artemis communications satellite, which relays the data to a ground station. The first successful transmission occurred in November 2001, with a data rate of 50 MBPS [ESA01]. As the beamwidth of the communications beam (~ 8 μrad at 0.83 μm) is much smaller than the RMS pointing error, pointing, acquisition and tracking (PAT) of the beam from one satellite to the other is a critical task [Tol98]. Currently, a high power laser beacon and two CCD cameras, one for acquisition, another for tracking, have to be employed to derive the error signals for the closed loop pointing system [Bai94]. It is believed that if a coherence-based approach is used, given the very high sensitivities demonstrated later in this thesis for the detection of a laser, a dedicated laser beacon would no longer be required, and a single laser beam could be employed both for PAT and communications.

The last application of the list of envisaged uses of coherence-based techniques in detection or imaging, related to the previous one, is the location and mapping of dim star positions at night for astronomy, or even comets. These celestial bodies possess chemical components that irradiate spectral spikes at selected wavelengths, having characteristic spectral signatures that are already used by conventional spectroscopic methods to identify them.

In conclusion, it is the author's belief that, besides the application of coherence-based techniques in the remote sensing of gases, previously explored by other workers, such techniques can be useful in several other applications, such as the detection of missiles and aircraft, biomedical specimens, chemicals in forensic science, and in inter-satellite communications. As will be shown in chapter two, previous works on coherence detection did not include intensive post-detector signal processing, which we exploit in this thesis. The examples given demonstrate that the use of coherence-based techniques for detection and imaging, principally of weak signals in heavy clutter, may yield better performance than existing intensity-based systems, and is, hence, worth investigating. This is the motivation for the research described in this thesis, which, although not application-oriented, intends to assess the feasibility of the proposed approach. The technique, partially invented by the author, and described in chapter three, was named in this thesis *"Interferogram Phase Step Shift"* (IPSS), and relies on measuring the displacement of a phase feature that occurs in an interferogram, which is proportional to the increase in the net degree of coherence of a scene.

1.2. Aims

The main aim of the research described in this thesis is to study, for the first time, an optical detection technique that uses the displacement of a phase step in an interferogram to detect a coherent emission or absorption in an incoherent background. In order to achieve this aim, the principal aims are:

I. To implement, in the laboratory, an optical detection system based on an interferometer to record the interferograms required by the study, using both existing and new designs;

II. To develop, optimise and test a method to determine the location of the phase step in the interferogram, and to demonstrate its effectiveness by comparison with other methods;

III. To derive a theory capable of predicting the displacement of the phase step as a function of target strength, for both coherent and partially coherent targets;

IV. To investigate the centre wavelength position and the bandwidth for the initial band pass filter for detection optimisation, and for a reasonable variety of targets;

V. To assess the sensitivity of the detection system developed by constructing its Receiver Operating Characteristic (ROC) curves;

VI. To compare the performance of the system developed with the performance of existing optical detection systems.

Additional aims are:

- To characterise the light sources used in the experiments, through measurements of coherence length;
- To develop variations of the technique that use no moving parts, by creating an image based on the target-to-background coherence contrast.

1.3. Layout

The thesis is organised in nine chapters. New material is presented from section 3.3 onwards. Chapter two presents a review of optical detection techniques usually employed in the detection of dim targets in highly cluttered backgrounds, based either on intensity, spectrum or coherence. Chapter three reviews, in the first two sections, the theoretical background required for proper understanding of the Interferogram Phase Step Shift (IPSS) technique. This theory is not new, but is included to support the advanced theory presented in chapters seven and eight. Section 3.3 describes the high level concepts of the IPSS technique, without detailing the new algorithm invented by the author, which is done in chapter seven. In chapter four a slight deviation of the principal aims of the thesis is allowed, to study the temporal coherence properties of the partially coherent light sources used in later experiments, through measurements of coherence length performed by recording long interferograms, similarly to what is done in Fourier transform spectroscopy (FTS). Interesting work on the characterisation of Resonant-Cavity LEDs produced at Imperial College is also presented there, which led to a published joint paper [Cou02].

Chapter five is the main experimental chapter of the thesis. It describes the experiments to detect coherent and partially coherent light sources using interferograms generated in the temporal domain. Three experimental arrangements are included, two using piezoelectric transducers, and one using a translation stage. The results, which depend on the algorithm used to retrieve the phase step (described in chapter seven), are depicted in this chapter. Additionally, the effects of the variation of the target-to-filter bandwidth ratio and central wavelength offset are assessed experimentally here. Chapter six, after displaying images of the interferograms acquired using a CCD camera, repeats some of the experiments of chapter five, this time producing interferograms in the spatial domain. In order to allow proper performance comparison, the chapter also includes the replication of imaging experiments from previous workers. A novel demonstration of coherence imaging using IPSS is included as well.

Chapter seven describes the development of the computer algorithm to extract the phase step from the recorded interferograms. The method for constructing the Receiver Operating Characteristic (ROC) curves is given, and the corresponding curves for IPSS, as well as for two competing techniques are calculated. A comparison between the three algorithms is then made using these curves. The detection performance of IPSS is then compared with the performance of other techniques through their probabilities of detection and false alarm. A novel, advanced theoretical model is presented in chapter eight. The model predicts the performance of the IPSS system under a set of assumptions. The predictions involve the shape of the shift vs. SCR curve, the effects of target-to filter bandwidth ratio, and central wavelength offset. A simulation tool that includes experimental effects not predictable using the theory is presented, and is used to predict the performance more accurately. The effects of polarisation on the SCR perceived by the detection system, as well as a comparison between experimental and theoretical results are included in this chapter. Finally, chapter nine summarises the conclusions from previous chapters, comments on the achievements and limitations of the research work in this thesis, discusses the advantages and disadvantages of the IPSS technique, and suggests directions for future work.

Chapter 2

Review of Relevant Optical Detection Techniques

This chapter presents a review of optical detection techniques, which are relevant for comparison with the approach under investigation. It starts with algorithms that detect targets using solely the intensity contrast between target and background, in a "single frame, single band" basis. The review proceeds describing techniques used in the detection of weak intensity targets, including the use of successive frames to explore temporal signature contrast and the "track-before-detect" approach. Then the use of more than one spectral band is introduced, in either dual-band or multispectral algorithms. In the next section, the emphasis is shifted from the intensity of the radiation to the spectral signature contrast, through the description of interferometric, spectroscopic techniques, highlighting the use of imaging spectrometers and Fourier Transform spectrometers. The fourth section of the chapter presents the detection techniques based on coherence, including optical coherence tomography. This section also presents the Optical Transform Image Modulation (OTIM) technique, from which Interferogram Phase Step Shift (IPSS) is derived, and various techniques based on it. Finally, a conclusion section summarises the detection sensitivities of the investigated methods in table 2.1 and presents trends in the area of weak target detection.

2.1. Intensity-based techniques

2.1.1. Introduction

The task of detecting a target embedded in a natural or man-made background is usually carried out using the spatial distribution of the intensity collected by each pixel in an imaging array as a discriminator. This is applicable both to visible and infrared wavelengths. In the visible band, detection and tracking are performed using the brightness contrast, which is a consequence of the different reflectivities of materials in that band. In this chapter, we will concentrate our review in the infrared. In this band, the intensity contrast appears as a combination of different temperatures and emissivities of the various objects in a scene, leading to the collection of different irradiances by different elements in the imaging array. The mechanisms of thermal radiative transfer and image formation are outside the scope of this thesis, and can be found in many textbooks [Zis93] [Hud69]. The most common infrared detection equipment to date are called FLIR (Forward-Looking Infrared) and IRST (Infrared Search and Track). The differences between them are well established in the literature [Cam93]. While FLIR is mostly concerned with forming and displaying an image, mainly for navigational purposes, IRST, as the name says, detects and tracks targets from the array of intensities, not necessarily forming an image. IRST systems can be employed either from the ground, ships or aircraft. They usually have a relatively small vertical coverage angle, not more than ten degrees, and scan 360° horizontally [Cam93]. In ships, their most important application is the detection of sea-skimming missiles, in addition to fire control radars, which suffer from fading effects due to multipath reflections at the sea surface [Cam93]. As IRST is mostly concerned with detection, while FLIR is with imaging, this section will concentrate in the former, describing some of its detection algorithms.

2.1.2. Single-frame, single-band processing

The simplest processing applied to an IR image in order to detect targets uses an IR detector or array of detectors with enough responsivity in one of the atmospheric windows, mainly 3-5 μm and 8-14 μm, and processes a single frame of two-dimensional intensities produced by this array at a time. These algorithms exploit the differences in spatial signature between target and background, thus being called spatial processing. As the detection system is designed to detect the target at ranges as far as possible, the solid angle subtended by the targets at these ranges is very small, when compared to the extension of background features such as clouds, terrain or sea, allowing the use of high-pass spatial filters for discrimination. These would reject the low spatial frequencies arising from clutter features, whilst passing target high spatial frequencies. Other less obvious spatial filtering techniques include edge detectors, median and Wiener filters, derived from the image-processing field [Wee98]. The next stage in the processing chain, shown in figure 2.1, is thresholding, where the intensity of all pixels in the image is compared with the same threshold to extract detection candidates. As this simple processor suffers from the variability of the clutter statistics across the field of view, one can think of a spatially adaptive threshold, after local estimations of clutter variance are performed in sub-regions of the image. For the same reasons, it is advantageous to apply this "local processing" concept to the spatial filtering as well. If we account for the detector array non-uniformities in responsivity and fixed pattern noise, a non-uniformity correction (NUC) has to be applied before spatial filtering. This algorithm is named "background normalisation" [Bax00], as it accounts, at least in theory, for the spatial variations of the background.

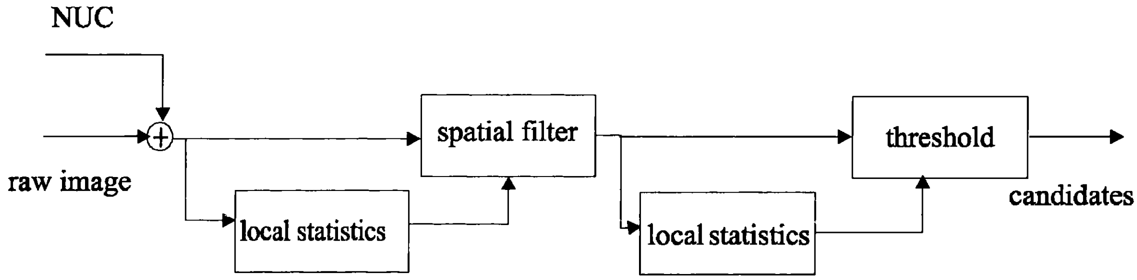


Fig. 2.1 - A typical spatial processing algorithm [Bax00]

2.1.3. Multiframe processors

Although spatial filtering can provide a baseline performance for clutter rejection, it suffers from the drawbacks of depending heavily on the scenario [Hil00] and of achieving a maximum clutter rejection ratio of 10 dB [Tar00]. Hilliard [Hil00] presented an extensive survey of clutter rejection algorithms, both spatial and temporal, applied to the detection of use of ordnance in the battlefield. The comparison of 21 different algorithms using clutter rejection metrics favoured a temporal Infinite Impulse Response (IIR) filter, with most temporal filters ranking among the first five. The limitations in spatial processing have justified the introduction of temporal discrimination in target detection algorithms. The thresholded, candidate targets coming from the spatial processor are correlated frame after frame, and if successive detections occur within a spatial window, a sub-part the Field of View (FOV), a target track is formed. The clutter rejection properties can be improved by taking the difference between a frame and its predecessor. Assuming the target is moving at a minimum speed of one pixel/frame, and that the background is stationary, then the differentiation will produce a target image no longer limited by clutter, but rather limited by temporal noise, and the target track can be recovered by simple thresholding. The performance of this approach, named velocity filtering [Cam93], is hampered by the need for proper registration between frames, to compensate for the effects of platform motion and sensor pointing jitter and drift. These problems are considered the major factor why IR scanning and staring sensors employ primarily spatial, rather than spatial-temporal, signal processors [Tar00]. In this work, Tartarovsky and Blazek state that line-of-sight (LOS) stabilisation alone cannot eliminate the distortions caused by LOS jitter, and propose an adaptive, non-linear, spatial-temporal filter. The coefficients of this filter are adaptively calculated using background parameter estimation. Using this algorithm and simulated data, the authors claim to have obtained a clutter rejection of 16.5 dB, in comparison with 6.2 dB obtained with spatial filtering only.

In a typical spatial-temporal approach, the pixels above the threshold, consequently detected, are registered in sequential frames to form a target track, using algorithms that are outside the scope of this thesis. These techniques are affected by rapid temporal variations caused by sun glint and aircraft motion [Pel97], and cannot cope with signal-to-noise ratios below 10 dB [Bra96]. When detection is declared via thresholding, the statistical information about that particular candidate is neglected, as the processor has to look for a next point to form the target track. A more efficient way of using the available information is to keep data about target positions and velocities stored before thresholding, declaring detection only after a number of frames is processed and a track is formed. This set of techniques is named Track-Before-Detect (TBD) [Wei95] [Ton96] [Mer97], made possible by the continuous increase in computing power, and representing the current state of the art in multiframe, single band target detection. A good review of early work in these techniques is given by Tonissen and Evans [Ton96]. In TBD a data cube is formed with one temporal and two spatial dimensions, to be processed by three-dimensional matched filters or dynamic programming algorithms. In the latter case [Ton96], the tracking problem (solved before the detection problem) is modelled as a state-space three-dimensional matrix where, from the current position of the target, the algorithm searches over all possible state sequences, returning the best sequence according to a merit function, or, alternatively, all the sequences exceeding a given threshold. The probabilities of detection and false alarm are determined by this threshold as usual, but processing gains arise from the integration of target energy throughout successive frames.

Another interesting work was published by Branlund, Davis and Lindgren [Bra96], where claims of simulated target detection with an SNR of 3 dB are made. The most important parts of the algorithm are image segmentation, spatial filtering, a Bayesian Field Tracker TBD and false alarm mitigation, using the likelihood ratio, a statistical concept found, for example, in Van Trees [Van68]. The spatial filtering consists of a two-dimensional matched filter, which convolves an estimate of the target spatial shape (for a sub-pixel target, the sensor point spread function) with the measured spatial intensity profile, corrected by the estimated mean and variance of the measurement (background). The Bayesian Field Tracker is, again, a state estimation machine where each update is governed by the likelihood of the presence of the target, calculated from the output of the same matched filter, assuming a composite Gaussian background distribution. Off-line processing of real backgrounds with simulated targets detected the target at a range of 23 km, corresponding to an SNR of 5 dB. Members of the same company published further work on this algorithm later [Mer97], where 34 frames in real data had to be integrated to declare the presence of an F1 Mirage aircraft in the image sequence. This long integration time (more than a second in typical 25 Hz imagers) not only represents a considerable reduction in the sensor update rate, but also gives an idea of the amount of required computing power to implement this sort of algorithm in real time. Wei, Zeidler and Ku [Wei95] maintain that state estimation approaches are not suitable for image processing, due to the amount of computing power required. In this work they also show that, for TBD to have better performance than single-frame techniques, a minimum SNR is required that depends on target velocity. This SNR is presented as being –3.7 dB for a velocity of 1 pixel/frame, and 2.2 dB for 2 pixels/frame. The stated numbers do not mean that targets will be detected at these SNR, but rather that, below these, the computing burden of TBD is unacceptable. This work [Wei95] mentions that a common operating point for electro-optical sensors would be a probability of detection of 0.9, and a probability of false alarm of 0.1. Using their model, they calculate that a single frame detection system would require a SNR of 8.2 dB to achieve these probabilities, and the integration of 4 frames at a velocity of 1 pixel/frame would require 6 dB to do the same.

More recently, Tartarovsky, Kligys and Petrov [Tar99] have recast the moving target detection problem as an abrupt change in the image pixel probabilistic distribution. They defend that current Bayesian or likelihood ratio detection techniques are not suitable for multitarget tracking, because they ignore the event of target disappearance. The paper proposes a sequential detection algorithm, in which a decision about target presence is made at each frame update. The hypothesis test is composed of three hypotheses: the target did not appear, appeared and is still present, or appeared and disappeared. The algorithm includes the optimisation of detection time delay, constrained by the probability of false alarm, and its detection statistics uses the estimates of target spatial location, based on optimal spatial-temporal nonlinear filtering. It is claimed that a simulated target with SNR = -6.6 dB was detected in real IR backgrounds, with a probability of false alarm of 1/60, after integration of 20 frames, leading to an output SNR of 4 dB. Although no statement about target velocity is made, this result appears better than the input SNR limit stated by Wei, Zeidler and Ku [Wei95]. This is also the lowest SNR detected by an intensity-based system, although with a penalty of a 20 second time delay.

An interesting exercise was performed [McC96] by the Wright Laboratory of the U.S. Air Force. The DICE (Detection in Clutter Enhancement) program included three different proposed solutions to assess the problem of detection of low observable targets. The basic requirements of the program were: for a target with irradiance intensity of 10^{-10} W/cm^2, a probability of detection of 0.9, probability of false alarm of less than one per hour, and declaration of threat 2 seconds after detection threshold crossing. An important conclusion of the study is that "major improvements in passive IR threat warning will come about only by incorporating new dimensionality (i.e., temporal, spectral or both) into the solution". TBD has been using the temporal domain, whilst multispectral techniques have introduced the use of the spectrum. The first of the three proposed solutions involved a spatial-temporal approach consisting of electronic frame stabilisation, adaptive spatial filtering and adaptive, multiframe temporal filtering, but the project [McC96] was terminated before completion. The second solution was the cascade of a velocity filter and a fuzzy, Bayesian system called MLANS (Maximum Likelihood Adaptive Neural System). The first stage is a clutter removal algorithm, based on a three-dimensional filter, whilst MLANS is, once more, a state estimation machine. The only results mentioned were the achievement of the requirements, not referring to SNR. The third approach described in the paper adapts an existing morphology-based non-linear signal processing technique named Geometric Whitening Filter to dual-band data, and will be described in the next section.

The last example in this sub-section is an algorithm proposed by Diani and others [Dia01]. It provides background removal in combined sea-sky cluttered scenes by assuming target and background lie in two orthogonal sub-spaces, and projecting each column of the image matrix onto the target sub-space, calculated from the background sub-space. The latter is estimated from the data, using the discrete Karhunen-Loève transform (DKLT), whose steps are: calculation of the image matrix basis vectors e_i by applying singular value decomposition to this matrix; calculation of the correlation matrix R_x by evaluating $R_x e_i = \lambda_i . e_i$, where i varies from one to the number of rows in the image matrix, and λ_i are the eigenvalues; and the extraction of the background sub-space by assuming it is represented by the eigenvectors that have the largest eigenvalues in R_x. Experimental results include a set of 16 frames, which are averaged to increase the SNR, and have an initial SCR of 0 dB. After application of the algorithm, the target is successfully detected, and an SCR gain of 5 dB is reported. Among the drawbacks of the technique are the computational complexity and the reliance on the background temporal correlation, whose problems where previously mentioned.

2.2. Dual band and multispectral techniques

2.2.1. Introduction

As mentioned in the first chapter of this thesis, the differences in spectral emissivity and temperature between targets and backgrounds make the use of more than one spectral band very attractive for use in detection systems, as a background suppressing technique. Multispectral techniques applied to detection came from the field of remote sensing, where they find use in atmospheric and geological studies. These techniques do not suffer from registration problems as much as spatial-temporal approaches do, although, on the other hand, some processing effort has to be spent on the removal of the effects of temperature [Cam93]. The most common approach is to treat the problem as a decision between multiple hypotheses, applying likelihood ratio criteria.

2.2.2. Dual band processors

Due to the presence of the blue spike [Cam93], mentioned in chapter one, in the spectral signature of missile plumes, dual band processors are being intensively used in missile warning sensors [McC96][Pel97][San96][Mon96][Mon00]. In the previous section we started reviewing solutions proposed in the DICE program, of which the only one not yet mentioned is the dual-band processor presented by Peli [McC96][Pel97]. It consists of two stages in tandem, the first removing clutter, and the second increasing the SNR. In the first stage, a primary spectral band has to be found, where the target has maximum signature and sun glints a small spectral density, possibly close to the CO_2 spike region (4.25 μm); the second or reference band needs a target with less contrast, and a larger solar glint peak, which usually happens at shorter wavelengths due to the sun's higher blackbody temperature. Subtracting the two bands will give a whitened image, where the background will be spectrally flatter, the target will display increased positive contrast, and sun glints a negative contrast. An estimation algorithm has to be included that extrapolates the background radiance in the reference band to the primary band, generating an estimated "target free" image. Peli argues that the linear estimation algorithm depends highly on clutter correlation in the two bands, and that it is computationally intensive. As a solution to this problem a Geometric Whitening Filter is presented, which consists of a series of morphological operations like dilation, erosion and edge enhancement. A proper combination of these operations achieves the desired level of inter-band correlation in each pixel, required by the whitening process. With the image now being dominated by noise and no longer by clutter, the second stage in the processing chain is the use of a track-before-detect approach, similar to the ones described in the previous section. Claimed results of such processing are a 100-fold reduction in false exceedance rates, or a 45% increase in target detection range, meeting the DICE program requirements. Receiver Operating Characteristic (ROC) curves are given, which are used for comparison in chapter seven.

The U.S. Air Force's Wright Laboratory has produced extensive research on the characteristics of missile signatures [San96][Mon96], in order to obtain an effective missile-warning sensor, capable of operating under heavy clutter conditions. With this characterisation data in hand, a study on dual-band algorithm comparison was performed [Bax00], and the best algorithm received further development [Mon00]. Baxley, Sanderson, Montgomery and McCalmont [Bax00] compared three single-frame algorithms: simple thresholding, taken as baseline and discussed in section 2.1, background normalisation (previously discussed as well) and two-colour correlated background, which is a linear background estimation model, reviewed previously [Pel97]. The conclusion of this study is that the simple threshold only works well in areas without structured clutter, the correlated background approach offers an intermediate performance, and, not surprisingly, the dual band algorithm gives the best performance. The former had higher probabilities of detection, when plotted against time, for a real IR sequence of 80000 frames. An interesting algorithm comparison plot is a scattergram where the probabilities of detection of the two algorithms being compared are plotted one against the other, for the same probabilities of false alarm and SNR. In a plot where the abscissa axis is either simple threshold or correlated background, and the dual band algorithm is in the ordinate axis, the sample points lie above the line crossing the origin at 45°, demonstrating its better performance.

Perhaps the most comprehensive and computationally heavy detection algorithm found in the literature is given by Montgomery, Sanderson and Baxley [Mon00]. Unlike most other examples, which tend to use pre-recorded or simulated images, in this missile-warning algorithm development a test bed with a real IR sensor was assembled. It comprises an f/2.1 optical telescope, a filter wheel to select the spectral band, a 256 x 256 InSb detector array, and a real-time processor with a 32 x 32 processor array, using parallel computing in a SIMD (single instruction, multiple data) architecture. The processor was configured to deliver a throughput of 31 8-bit GOPS and 0.6 GFLOPS, about 25% of its full capacity. Internal communication processor-memory is done at 3.8 GB/s. The use of such a supercomputer (denomination given by the author) is justified by the size of the spatial, temporal and spectral algorithm, whose flowchart is depicted in figure 2.2. The blue and red channels (explained in more detail later) receive non-uniformity corrections consisting of spatially varying gain and offset; frame registration is accomplished with a gradient-compensation technique, to cancel platform motion; then spatial filtering is achieved with highpass filtering, and subtraction of a median-filtered background estimate from the image. The next steps are a temporal, velocity filter, which assumes target and background move at different speeds, and a dual band, colour-ratio filter. For the temporal stage to operate properly a certain degree of background temporal correlation is required, which is achieved through high frame rates (100 Hz).

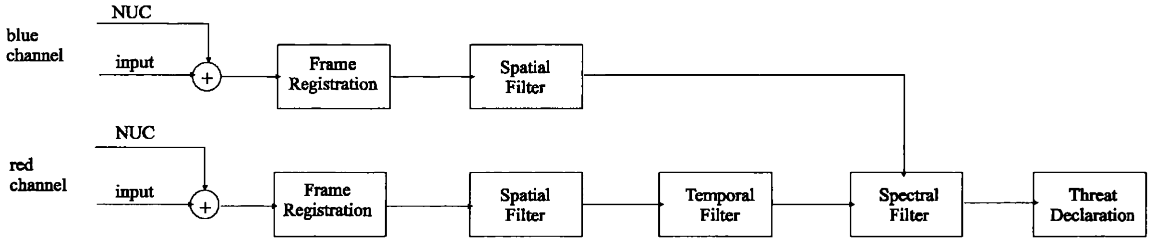


Fig. 2.2 - Missile warning algorithm by Montgomery, Sanderson and Baxley [Mon00]

The spectral discrimination is achieved exploring the target-to-background signature contrast in two bands: 3.52-4.03 μm and 4.57-4.71 μm, and works similarly to the work by Peli [Pel97], but employing a ratio rather than a subtraction between the intensities in the two bands. The ratio data is then thresholded to be handed to the last stage. It is claimed that this spectral filter works with the high-intensity values (outliers) in the background distribution, where most of the false detections occur. The final stage in the processing chain is threat declaration. Even with four-dimensional discrimination (two-dimensional spatial, temporal and spectral), temporally correlated noise coming from electronics and other incidental targets still degrades the performance. This stage is implemented as a Kalman filter [Ope89], which receives information from the platform itself and target motion, and forms tracks with the detections from the previous stage. This module was still under development by the authors at the time of publishing.

The algorithm is computationally very intensive. It requires in between 1500 and 3500 operations per pixel excluding the threat declaration stage; around 100,000 clock cycles per frame, some 16.10^6 clock cycles/sec. With this processor power, up to 28 targets can be passed to the platform countermeasures system. Results are presented from real missile firings, with platforms at helicopter speed travelling at 7500 ft altitude. No absolute input SNR is given, but the algorithm is capable of raising the normalised raw image SNR by 11.2 dB. Incorporating into this figure of merit the reduction in false positives, a claim of a total SNR gain of 52 dB is made, although the gain calculation method is not clear.

It is worth offering a few comments about this work ([Mon00]), due to its high SNR gain figure. It has put together most of the known clutter rejection techniques, and placed them in tandem. Because of this, algorithm complexity grew proportionally, requiring the use of a "massively parallel supercomputer" (quoted from the paper). The relative results quoted do not include probabilities of detection and of false alarms, although a 100-fold in the latter is claimed. Some of the assumptions are not very realistic. It is assumed that the target intensity is additive to that of the background, being targets inserted by "cut and paste". As an example, in the case of an SCR of 5, presumably within the sensor dynamic range, this would represent a 20% error in combined intensity. Furthermore, effects of target shadowing the background need to be considered. This procedure requires some sophisticated processing, as argued by Hilliard [Hil00], who developed a special code especially for target insertion.

2.2.3. Multispectral processors

In situations where thermal contrast is poor and target signatures are not easily discernible using dual band techniques, more than two spectral bands may be used to increase target discrimination against backgrounds. Multispectral techniques [Cam93] are widely used in remote sensing and defence applications. Their aim is to increase target detectability through exploration of differences in spectral emissivity, rather than temperature. Vibrational and rotational oscillations of molecules in the atmosphere, soil, vegetation and man-made objects create absorption and re-emission features that characterise these bodies as selective radiators [Hud69]. The combined effects of these features with the blackbody emission characteristics of objects at different temperatures generate spectral signatures that allow discrimination in wider wavelength bands than it is achievable with two bands only.

In section 2.1, the concept of a data cube was introduced to aid visualisation of data to be processed by detection and tracking algorithms. As a new dimension is included as an output of the sensor (spectrum), a four-dimensional space is created, where several three-dimensional data cubes would exist, one for each band. In order to collect such data, an imaging sensor capable of recording one temporal and two spatial dimensions has to be combined with a spectrally selective device, which can be either an optical filter wheel, a dispersive prism or grating, or an interferometric Fourier transform-based device. Whilst the latter two are discussed in the next section, a good example of the first method is given by Ljunberg and others [Lju97]. Unlike the majority of remote sensing applications, defence tasks usually have to be carried out in real time, posing a technological challenge to the generation and processing of images. The advances in computing and in detector array processing, with higher frame rates, made these tasks possible. In this example, a 128 x 128 Mercury-Cadmium Telluride (MCT) array with a frame rate of 200 Hz, covering the 2-5 μm band, is combined with six 500 nm wide filters mounted on a rotating wheel. An industrial computer synchronises the motor that rotates the filter wheel and the array readout electronics, such that six spectral images can be sequentially generated with approximately 33 Hz frame rate each. The amount of memory available in the computer allowed the storage of only 20 seconds of images, illustrating the difficulties with handling the large amounts of data produced by this kind of sensor. Due to the relatively wide width of each band (500 nm), the paper just presents qualitative remarks about visible band-to-band differences, such as poor imaging performance in the 2.5-3.0 μm band, due to sun glint, good transmissivity in the 3.5-4.0 μm band, and higher contrast of an aircraft in the 4.0-4.5 μm band, due to the CO_2 spike phenomenon. No detection algorithms are mentioned in this work.

Cheung and others [Che97] present the multispectral detection problem more rigorously, as a composite hypothesis-testing problem with:

$$\mathrm{H_0}:\ \bar{x} = \begin{bmatrix} s_{b1}(\varepsilon,T) \\ \cdot \\ \cdot \\ s_{bN}(\varepsilon,T) \end{bmatrix} + \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_N \end{bmatrix} \text{ versus } \mathrm{H_1}:\ \bar{x} = \begin{bmatrix} s_{t1}(\varepsilon,T) \\ \cdot \\ \cdot \\ s_{tN}(\varepsilon,T) \end{bmatrix} + \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_N \end{bmatrix} \qquad (2.1),$$

where $\bar{x} = [x_1 .....x_N]^T$ is the vector of energies observed in each one of the N bands, s is the energy component in each band, with subscript b for background and t for target, ε is the emissivity, T is the temperature, and w is the noise of the detection system in each band. A target, when it exists, is assumed to fill a whole pixel. Under the assumptions that the multispectral signal and the noise are normally distributed, a sufficient test statistic [Van68] is generated by calculating the difference between a target-absent and a target-present statistic, represented by a spectrally matched filter (SMF). This difference is then compared to a threshold to declare detection, being the receiver performance defined by the threshold level. The SMF concept is widely used in multispectral processing [Che97] [Sch96] [Sin98] [Sin99] [Sin00] [Shn00], and is an extension of the standard matched filter [McD98] to a multiband system. It can be represented by:

$\bar{F} = \bar{X}^T.C^{-1}$, where F is the filter row vector, with one coefficient for each band, C is the background covariance matrix, and $\bar{X}^T$ is the transposed input vector $\bar{s}$ in eq.2.1.

Without detailing the mathematical formalism, which is outside the scope of this thesis, Cheung and the co-authors claimed, based on simulations with greyscale images, that this approach suppresses clutter and whitens the multispectral data for thresholding, although no detection performance figures are given.

A problem with hypothesis-testing-based multispectral processing is the difficulty in separating emissivity from temperature contributions to the detected radiance value. If the receiver is to operate on differences in emissivity across different bands, the effect of different objects having different temperatures has to be accounted for in the detection process. Many workers did recognise this issue [Che97] [Lju97]. Cheung and others [Che97] point out that inaccuracies in temperature estimation may lead to severe losses in detection performance. Ljungberg and others [Lju97] used a temperature calibration procedure.

Singer and co-authors published a series of papers [Sin98] [Sin99] [Sin00] where a space-spectral matched filter is introduced. Like Ljunberg and others [Lju97], they also highlight the use of modern, high-speed focal plane arrays in the acquisition of multispectral data. As target spectral signatures are affected by the environment and, thus, may not be known, instead of comparing the measured signature with a spectral reference a statistical test is performed to determine if this signature is sufficiently different from the background's signature. The detection system is, thence, composed of a matched filter rejecting clutter, and a spectral anomaly detector designed to circumvent losses caused by lack of a priori knowledge. One of their findings is that, from two bands onwards, additional bands do not increase the signal-to-clutter ratio (SCR), but do increase the false alarm rate. This is due to the increase in the number of false spectral features detected by the anomaly detector, as more bands are included. Using the track features concept [Sin99], the algorithm without a priori knowledge of target spectrum presents the same performance as an algorithm that uses such knowledge, the authors claim. The track features concept is similar to TBD, as additional information about the target, such as the log-likelihood ratio [Van68], is retained after use by the anomaly detector in order to form a spectral information track. In [Sin98], six bands from 8.2 to 11.2 μm are employed, the narrowest ones encompassing 400 nm. SCR gains of 15.5 dB, in opposition to 12.3 dB obtained by single-band, spatial-temporal processing, are referenced. ROC curves are presented, which will be used for comparison purposes in chapter seven.

A not so recent, but comprehensive work on multispectral detection was presented by Schwartz and others [Sch96]. Using spectral signatures of military targets and natural backgrounds, recorded using Fourier Transform Spectrometers (described in the next section), the authors explore the spectral matched filter concept already described to characterise detection performance, and perform band optimisation procedures that will maximise that performance. The aim is to choose adequate band locations to explore target spectral features, in this case assumed to be known a priori. Unlike single band systems, which tend to be clutter-limited, the performance of multispectral systems is usually limited by noise generated by the sensor itself [Sch96]. As the increase in the number of bands increases system noise, background correlation tends to degrade. Tight maximum system noise requirements are posed because of this fact when using three or more bands, to retain multispectral clutter rejection advantages. Although causing an increase in SCR, increasing the number of bands may degrade the performance if system noise is not controlled, as it can cause estimation errors in the background covariance matrix. This fact was possibly why Singer [Sin98] found out that three bands were worse than two. Very high band-to-band background correlation was found by Schwarz in the LWIR band (8-12 μm), which is used for clutter whitening. The approach has problems in the MWIR (3-5 μm), due to low background correlation, caused by the presence of water vapour and carbon dioxide absorption bands in this band. Performance is gauged by the frequency of occurrence of target/background pairs with SCR higher than 8 dB, considered an indication of good detection performance. An SCR gain of 8 dB is achieved when using two bands, 17 dB using three, and 21 dB when using four. ROC curves are given, which will be used in chapter seven.

A last example of multispectral approaches was presented by Shnitser and others [Shn00], and illustrates the use of other properties of light in target contrast enhancement. The development combined multispectral with polarisation discrimination, doubling the number of bands in the multispectral system without increasing system noise. The embodiment, depicted in figure 2.3, is based in software-controlled acousto-optic tuneable filter and liquid crystal achromatic rotator, and generated non-video rate imagery (4 frames per second). Images are recorded in a CCD camera and digitised in a frame grabber. A PC controls the digitiser and the optical device drivers.

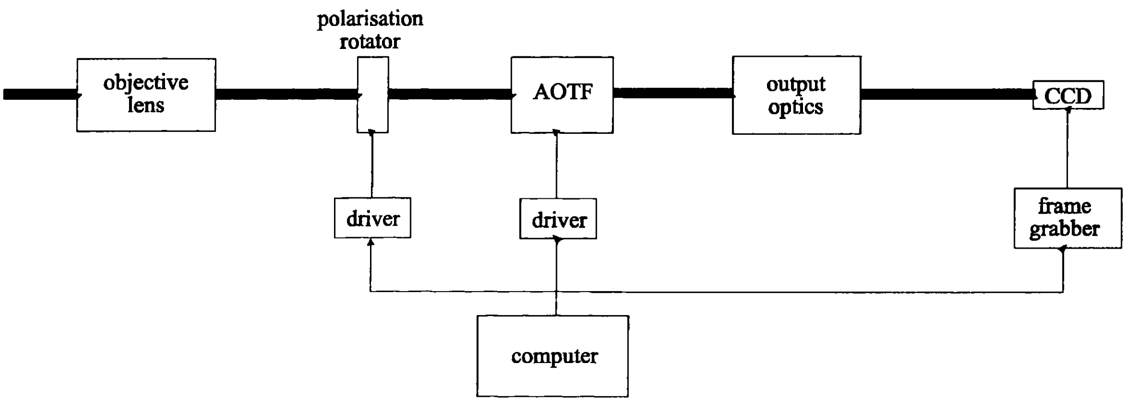


Fig. 2.3 - Spectro-polarimetric imaging system [Shn00]. Thick lines indicate the light path.

The experimental demonstrator worked in the visible band. The input optics image the scene into the liquid crystal, used to obtain 90 degrees polarisation rotation with a performance independent of wavelength. The polariser is controlled by a voltage that, when absent, allows the rotation, and, when present, blocks it. This rotation explores the birefringence properties of the twisted nematic material [Hec98]. Hence, by changing the control voltage in an on-off fashion, two different polarisation states can be admitted to the next stage, which is an acousto-optic tuneable filter. This filter provides spectral discrimination, once more based on the spectrally matched filter concept. Seven frames are required to generate one contrast-enhanced image, for each spectral band. When using a single band, which is the case of this demonstration, two frames need to be recorded in vertical polarisation, and two in the horizontal, due to the need of switching between positive and negative coefficients of the matched filter. Three more frames are required because of the relatively slow response time of the liquid crystal and the tuneable filter, quoted as a few tens of milliseconds. The demonstration comprised the discrimination of a piece of glass in simulated substrates based on polarisation contrast. Hardware limitations imposed restrictions to the significance of the results, such as the low frame rate (4 Hz), due to the low CCD camera frame rate (30 Hz) and low response time of the optical devices, and the use of a single spectral band. Although no sensitivity or SCR gain results are quoted, the work has the merit of demonstrating the possibility of combining two different dimensions (spectrum and polarisation) in a single detection system. Once more, the penalty is the large volume of data generated, requiring heavy processing resources. Another interesting feature is the use of pre-detector optical processing, which is also explored in IPSS.

In conclusion, the multispectral detection techniques presented in this sub-section can provide considerable SCR gains. However, assuming absence of a priori knowledge of target spectra causes heavy penalties to detection performance. The overall bands used in these systems are very wide [Lju97] [Sin98] [Sch96], leading to sub-bands reaching from 400 to 1000 nm. The detectors integrate the energy within these sub-bands, averaging out important spectral features caused by molecular absorption phenomena, which display much narrower bands [Hud69]. By disregarding target a-priori knowledge, the large bandwidths required can also generate large amounts of data, requiring more processing power and slowing down detection algorithms. These two problems are related to an argument to be presented when discussing the advantages of the IPSS technique, in chapter nine: if one knows what to look for, one can find it a lot faster. A third problem arises from the use of large imaging arrays, where non-uniformities and non-linearities between the array pixels degrade the background correlation that most techniques rely upon [Eis97].

2.3. Spectroscopic techniques

2.3.1. Introduction

In this section, we depart from the field of detection techniques to the one of spectroscopy. In the former, the emphasis is on creating algorithms capable of combining characteristics of the target in a small set of wide bands, to enhance its detectability. In the latter, the number of bands and their resolution is increased, but the main interest is in measuring an object's spectral signature, either with research purposes or to support the detection studies. As a consequence of this different emphasis, few researchers present sensitivity measures of their spectroscopic approaches, but rather the spectra they have obtained.

The instruments that make use of the spectral dimension were classified by Breckinridge [Bre96] according to the fractional bandwidth of the individual bands ($\Delta\lambda/\lambda$) into multispectral, when this ratio is about 0.1, hyperspectral, when it is about 0.01, or ultraspectral, when it is about 0.001. The first type was discussed in the last section, the second and third are discussed here. The term *imaging spectrometers* is used throughout the text, denominating instruments capable of obtaining data which is both spectrally and spatially resolved [Wol97], and which were the result of the addition of imaging capability to existing spectrometric techniques. Multispectral and hyperspectral instruments can be implemented with filtering techniques, either by rotating filter wheels [Lju97] or using tuneable filters [Mar99]. The very high resolution required in ultraspectral instruments can be achieved through dispersive spectrometers using prisms or gratings, or through Fourier-transform spectrometers. Because of their relevance as one of the foundations of the IPSS approach, only Fourier-transform spectrometers are discussed in this review. On the subject of imaging spectrometers in general, the interested reader is referred to review papers such as the ones by Persky [Per95] or Goetz [Goe95].

2.3.2. Fourier-transform spectrometers (FTS)

The operating principle of FTS is not as straightforward as happens in the direct spectroscopic methods, in which the spectral domain is directly measured by slicing the wavelength dimension in spectral bins. FTS is mathematically based on the Wiener-Khintchine theorem, found in many optics textbooks [Bor99] [Hec98], which states that the Fourier-transform of the autocorrelation function of a wave field is its power spectral density, or the squared modulus of its spectrum. This theorem is presented in more detail in chapter three, as well as other important theoretical aspects of FTS that serve as a basis for the IPSS technique. Based on the Wiener-Khintchine theorem, by measuring the autocorrelation function and computing a Fourier-transform, one could obtain an estimate of the spectrum. Although this possibility was known to Michelson and Rayleigh in the 19th century, it was only in the middle of the 20th

century (1948-1951) that Fellgett obtained what were probably the first spectra derived by Fourier transformation [Fel84], calculated manually. FTS became widely used only after computing resources were widespread, and, most importantly, after the invention of the Fast Fourier Transform (FFT) algorithm by Cooley and Tukey in 1965.

In spectroscopy of electromagnetic waves in the optical part of the spectrum, the principal device used to measure the autocorrelation function, or the self-coherence function in the optical jargon, is the Michelson Interferometer [Bor99] [Hec98], depicted in figure 2.4.

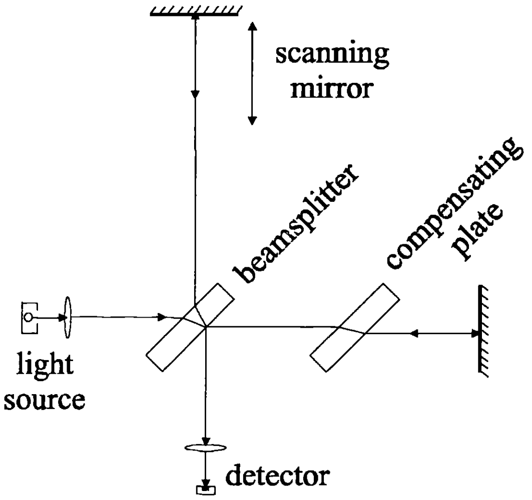


Fig. 2.4 – Michelson interferometer

In this instrument, light is admitted through optics that collimate the beam and increase the instrument's collection angle. The radiation reaches a beamsplitter, a plate with a finite thickness and a coating at the rear surface that transmits part of the radiation and reflects another part. The reflected and transmitted beams reach two mirrors, whose distances to the beamsplitter coated surface are very precisely controlled. The radiation reflected from the two mirrors comes back to the beamsplitter, where they recombine, interfere and exit the instrument via another set of optics, which focus the collimated beam into a small-area detector. While one mirror is longitudinally fixed, but is adjustable in tilt for alignment purposes, the other one is scanned longitudinally in order to vary the path difference. Because of the double path between the beamsplitter and the mirrors, the optical path difference between the two beams is twice the difference in mirror-beamsplitter distances. By scanning the moving mirror continuously, it is possible to measure successive maxima and minima of constructive and destructive interference of light, which appear as a carrier modulated by its self-coherence function. A compensating plate, made of the same bulk material as the beamsplitter, is used in one of the interferometer's arms to equalise the paths between the two beams irrespective of wavelength, hence making it broadband. An FTS instrument usually consists of an interferometer, and a computer which controls the scanning, performs the required FFT and applies many corrections to the data in order to remove measurement artefacts. Some of these corrections are discussed in chapter four.

When compared to dispersive instruments, interferometer-based spectrometers display two main advantages, named after their discoverers: the throughput or "Jacquinot advantage" [Cha79] relates to the much higher light-collecting capacity of the FTS instruments, when compared to spectrometers relying on narrow slits. As the throughput or *étendue* of an instrument is defined as the product of the detector area and its light-collecting solid angle, an optical invariant throughout the optical system [Cha79], Jacquinot has calculated a considerable advantage of FTS over dispersive instruments, which employ slits to select the desired part of the spectrum. The multiplex or "Fellgett advantage" [Cha79] is due to the fact that, in dispersive instruments, a single band is presented for integration in the detector at a time, whilst in FTS all wavelengths within the detector band are integrated during the whole measurement time. These advantages, allied to the very high spectral resolution achievable, made FTS the technique of choice in infrared spectroscopy, being these instruments named FTIR (Fourier Transform Infrared) spectrometers.

Nevertheless, some caveats about the FTS advantages have to be considered, as their existence in a particular instrument depends on its design and application. The multiplex advantage is only available when the noise is independent of signal power [Cha79], which happens whenever the sensitivity is either limited by background power, or by detector noise. If the noise increases with signal level, increasing detector integration time as happens with FTS will increase both signal and noise, not bringing any advantage. Fortunately, the occurrence of signal-dependent noise is relatively rare [Der96]. Another point is concerned with spatial resolution. When converted into imaging spectrometers, the throughput advantage is reduced as the number of pixels in the array increases [Pri97] [Des96] [Hor96]. This is simply due to the fact that, in order to achieve better imaging spatial resolution, the instantaneous field of view (IFOV) of each pixel has to be minute, eroding the wide FOV that otherwise would yield the throughput advantage.

Eismann and others [Eis96] present a comparison of four different spectroscopic imaging techniques: the use of filters, dispersive and two modes of FTS: the temporal mode, already described, and a spatial mode, where path difference is obtained by tilting one of the mirrors to obtain line (Fizeau) fringes [Hec98], which are recorded spatially in the detector array. The spatial mode has the advantage of using no moving parts, being more compact and robust, but at the expense of sensitivity due to the dispersion of energy among the detectors in one dimension. Due to this, the spatial FTS had the highest Noise Equivalent Spectral Radiance (NESR), which is the scene radiance change leading to an SNR of one [Per95], of all four designs. As the studied system was background-limited, the temporal FTS did not have the multiplex advantage over the filtered design (as shown in the previous paragraph). Comparing temporal FTS against dispersive, the latter had a better performance for high spectral resolutions, whilst the former performed better for relaxed resolutions, in agreement with the findings of Pritt and others [Pri97]. The temporal FTS design was the most immune to array non-uniformity and non-linearity, as even with high influence of these parameters a clutter suppression factor of 30 dB is still obtained [Eis97]. Horton [Hor96] presents an interferometer optical design with no moving parts to generate interference fringes, while keeping part of the multiplex benefits. It forms an autocorrelation function data cube by tilting one of the interferometer mirrors in one dimension and performing a lateral scan of the image in the other. Chapter six presents experiments with an imaging device that forms a similar data cube, but performing a longitudinal scan instead.

Another static FTS is presented by Ivanov [Iva00] that aims to circumvent one important limitation of spatially modulated FTS: the resolving power ($k_{\max} / \Delta k$) vs. free spectral range ($(k_{\max} - k_{\min}) / k_{\max}$) product is limited by the size of the array. (k_{max} is the maximum wavenumber, k_{min} is the minimum, and Δk is the spectral resolution). Increasing the mirror tilt angle enlarges the fringes. As one fringe corresponds to one wavelength, the path difference range is reduced and, consequently, the resolution is degraded, as both are proportional [Cha79]. The proposed solution is tilting one of the interferometer's mirrors in one direction, and giving it a stepped shape in the other, producing a folded interferogram with as many sections as the number of steps, much like the raster scan pattern in a TV set. Besides requiring very careful optical design of the imaging optics in order to avoid cross-talk effects between interferogram sections imaged by neighbouring mirror steps, the approach comprises a long image processing chain to compensate for dirt, aberrations, illumination non-uniformities, and the FFT. The method, demonstrated in the visible band, showed the possibility of using static FTS while maintaining resolving power and the throughput advantage, making this sort of instrument attractive for use in the UV, visible and NIR bands.

The temporal FTS method presents a trade-off because of the spectral resolution vs. path difference scan range already mentioned. An increase in resolution means a longer scan, more data generated and a longer FFT to be performed. Chern and Chao [Che96] presented a method that replaces the FFT with a spectral estimation method based on Principal Component Analysis [Kay88], and requires much less samples to estimate a spectrum from an interferogram. Although the experiment described in the paper only comprised the measurement of a single wavelength of laser radiation, by using 60 of the 1000 interferogram samples available, it was possible to measure the wavelength with an error of 3.10^{-4} at a simulated SNR of –3 dB. In the corresponding FFT-generated spectrum, the signal wavelength is not discernible. The avoidance of the FFT and the use of a smaller number of interferogram samples are the main characteristics of this method, which are shared by the IPSS method proposed in this thesis. In both cases, the reduction of the data volume is made possible by the use of a priori knowledge about the signal to be detected.

Thériault [Thé01] presented another example of use of a priori knowledge to reduce signal processing complexity, applied to the remote detection of gas clouds. The work concerns the practical demonstration of a differential detection method applied to FTS. It relies on the principle of background spectral subtraction to reduce false alarms. The FTS has two telescopes with a 5° angular offset, one pointed to the gas cloud, and the other to the background close to it, whose spectral radiances are subtracted in real time, prior to the application of a detection algorithm. A Simplex optimisation algorithm [Pre92] is used to adjust parameters required for deriving a calculated spectrum. The concentration-path length product, the most common figure-of-merit for gas sensors, is found by minimising the difference between the calculated and measured spectra. Obviously, in the estimation of the theoretical spectrum, the species of gas to be detected has to be known a priori. The detection of the gas cloud is declared using the correlation index between measured and calculated spectra as a parameter. The only numerical figure given as for sensitivity is 258 ppm-m, for ammonia, at a distance of 105 m. The critical issues in the use of the technique, according to the author, are the estimation of the gas cloud temperature, its influence on the spectral absorption coefficients used to generate an estimated spectrum, and concentration temporal fluctuations. Additionally, the spatial offset between the fields of view of the two telescopes employed in the instrument can produce errors, if the background spatial and temporal signatures are significantly different in these two FOVs. The author states that the remote sensing of gases using FTIR "has not been fully mastered yet".

A last example of FTS included in this review is the conversion of a commercial FTIR instrument in an imaging spectrometer [Nel00] for spectral characterisation of static images. The imaging capability is obtained by adding laterally scanning optics and a PC to control the longitudinally scanning mirrors and the spectrometer. The time taken to acquire an image depends on the spectral and spatial resolutions and on the SNR, which requires a variable number of scans to be averaged, and ranges from 3 seconds to 25 minutes. The spectrometer itself is faster, acquiring 65 spectra/sec with a resolution of 16 cm^{-1}, being the long acquisition time mentioned due to the very large amount of data generated by the instrument. The instrument also requires self-emission calibration to be performed every 4 hours. The paper presents good sensitivity figures, according to its author, namely an Noise Equivalent Temperature Difference (NETD) at 11 μm of 300 mK by integrating 16 scans, and 1 K when using 2 scans. The Noise Equivalent NESR is parameterised by resolution and number of scans, and a representative number is $3.2.10^{-7}$ $W/cm^2sr.cm^{-1}$ (or 3.2 μflicks) at 1 cm^{-1} resolution and 1 scan. An experiment of remote detection of ammonia is presented, but no sensitivity figures are given. Even with a low maximum spatial resolution (32x32 pixels), the system can generate amounts of data that are difficult to process for real time applications, which limits the application of the instrument to research purposes.

2.4. Coherence-based techniques

2.4.1. Introduction

As will be presented in the next chapter, a light wave possesses auto-correlation properties that can be used as discriminators in the detection process. Such properties are the spectrum, whose exploration was described previously, coherence, reviewed in this section, and polarisation, which is not part of this review, but has been increasingly used to enhance target-to-background contrast.

Coherence of light and its properties find a number of applications in many different areas of science. In this review, particularly, we will concentrate on its application to detection systems, overlooking, for example, its uses in coherence modulators used in telecommunications and in coherence-multiplexed fibre sensor systems [Kim99].

2.4.2. O.T.I.M.

OTIM stands for Optical Transform Image Modulation, a set of techniques invented by French [Fre72] designed to increase the contrast between targets or other man-made objects and a usually brighter background. Although its main features are presented in this section, the interested reader can probe further in a fairly comprehensive investigation of OTIM given by Sutton in his doctoral dissertation [Sut82].

As shown in chapter one, conventional electro-optical sensors rely on intensity differences to perform detection of objects against a background. This process is normally aided by spectral filtering of the input radiation, which limits the noise and clutter bandwidths to that of the filter. In order to do that, there must be prior knowledge of the desired signal wavelength, to maintain it inside the filtered band. The filtering approach may not solve the detection problem on its own, for even within the filtered spectral band, the background radiation may still be comparable to or higher than that from the desired signal. The filtered signal is then converted to an electrical signal in a photodetector, and as this detection process is incoherent, using only the signal intensity, the light to current conversion discards many other sources of information featured in the light wave, such as its phase, coherence, polarisation and spectrum [Fre84].

The OTIM approach aims to explore these other features of the radiation, in general, concentrating on coherence and spectrum, in order to extract the target signal from the background, despite its dim intensity. A processing gain is achieved by a selective modulation of the signal, as happens with communication systems, the difference being that this modulation is happening at the receiver, before the destructive and irreversible detection happens. Consequently, OTIM is a pre-detector technique. Target-to-background discrimination is achieved by exploring the coherence domain, where signal and background have different signatures, even if they have comparable radiated power. This coherence contrast is explained later in this section.

If it is assumed that the background is not affected by the modulation, which is achievable by careful design [Sut82], then the problem becomes one of ordinary demodulation, with signal and noise only, being background clutter totally rejected at the demodulator. Although the technique was focused in the visible and IR bands, it can be used in any spectral band where components are available, and can be used in addition to conventional post-detector techniques [Sut84], achieving further discrimination. Generally speaking, the technique can operate either in a spatial or in a temporal coherence mode. Two essential components are seen in all OTIM embodiments: the first is an interferometric device creating an interference fringe pattern, in the temporal mode, or a diffractive component generating a diffraction pattern, in the spatial mode. The second is a modulating device, to modulate the pattern generated in the first component. Background features, being incoherent, will not produce interference or diffraction, thus not being modulated. The possible uses of OTIM, discussed in this review, are temporal coherence filtering and target location. These, along with other possible uses, such as spatial coherence filtering, spectral profile measurement and coherence length measurement, are described by Sutton [Sut82].

2.4.2.1. Temporal coherence filtering

This subset of OTIM consists in discriminating the target from the background using the temporal coherence features that a target might have. It was mentioned in chapter one that man-made objects tend to have narrow spectral features within their spectral signatures, which may come from atomic emissions (solids and liquids), molecular rotational and vibrational emissions (gases), emission or reflection of coherent light (e.g. laser beams), among other mechanisms. As will be shown in chapter three, the narrow feature in a certain band of the spectral signature gives the target's radiation a higher coherence length, provided that the feature lies within the optical band of the sensor. Unlike the target, the background, although sometimes containing either high brightness or features in the spatial and temporal domains, usually has a smooth spectrum, resulting from the combination of radiation reflected from the sun or self emitted, both given by Planck's blackbody equation [Hud69].

As mentioned previously, there must be an interferometric device and a modulating one. Two basic arrangements are described in [Sut82]. In the first one, shown in figure 2.5, interference is provided by a Fresnel biprism, and the modulation is achieved by scanning a mirror. In the second (figure 2.6), the interferometric device is a Michelson interferometer, and an oscillating glass plate placed in one of the optical paths provides signal modulation.

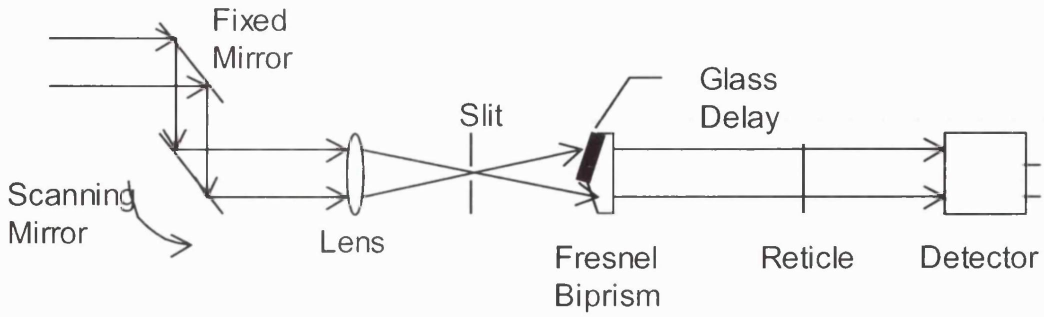


Fig. 2.5 – Temporal coherence filter based on a Fresnel biprism

In figure 2.5, light is admitted (through an input aperture) to the set of mirrors, one of them scanning the scene to be measured. The lens is used only to reduce the divergence of the light. The Fresnel biprism contains two slightly tilted input surfaces (around 2 degrees, exaggerated in the figure). This tilt will produce on the interference plane two partially overlapping regions, coming from beams travelling in the upper and lower halves. As the upper beam will also have to travel through the glass plate, there will be a delay given by:

$$\delta L = t(n_{glass} - n_{air}) \qquad (2.2),$$

where t is the thickness of the glass plate and n is the refractive index. The two beams recombine at the reticle plane, producing interference. The reticle has a geometry that matches the interference pattern. Modulation comes from the displacement of the input scene, originated from the mirror scan, through the slit, sampling different regions of this scene with time. If a target with a coherence length greater than δL comes into the field of view of the sensor, fringes will be produced at the reticle, and the relative displacement will produce spatial modulation at the detector. Background radiation will not interfere, producing a bias level that, after conversion to an electrical signal, can be removed by analogue or digital filtering.

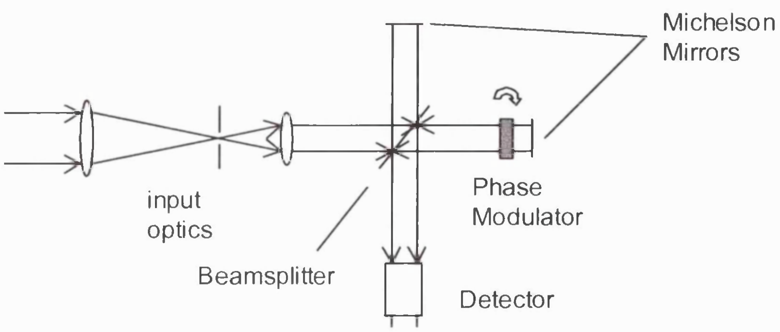


Fig. 2.6 – Temporal coherence filter based on a Michelson interferometer

In another version, the interference is provided by a Michelson Interferometer. As can be seen in figure 2.6, the entrance has a bi-convex lens. The beamsplitter splits and recombines the two beams, which interfere at the detector plane. High pass coherence filtering is achieved by introducing a bias delay in one of the interferometer's mirrors, so that only light with coherence lengths longer than that delay would produce fringes. Such an effect is detailed in chapter three, where the definition of coherence length is presented. Fringe modulation is achieved by twisting a glass plate placed in one of the optical paths, varying the plate thickness crossed by the beam, thus giving it a variable delay. As mentioned before, one of the essential features of this technique is fringe modulation, which can be implemented in many other ways. Any device capable of introducing a variable and controllable delay is suitable. In [Sut84] this is done by a pellicle beamsplitter whose longitudinal position is modulated by a sound wave from a loudspeaker. Electro-optic, acousto-optic and birefringent cells, with less reliability and alignment problems than the twisting scheme, may be considered as well.

In some cases the sought after signals might not be the higher degree of coherence ones, but rather some range of values. Then it is worth thinking of bandpass coherence filtering, whose aim is to produce interference fringes only from light with coherence lengths within that range. To understand how this can be achieved, it is necessary to consider two high pass systems, one with a delay L_1, and the other with a delay L_2. The derivation that follows is adapted from Sutton [Sut89]. Referring to figure 2.7, the fringe visibility is given by:

$$V(\tau) = \frac{I_{\max}(\tau) - I_{\min}(\tau)}{I_{\max}(\tau) + I_{\min}(\tau)}$$

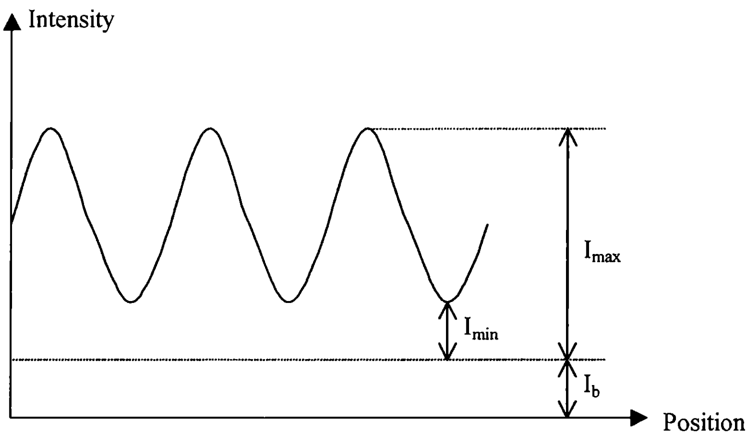


Figure 2.7 – Theoretical fringe intensity profile (sinusoid exaggerated) [Sut82]

The intensities measured by a detector will vary between a maximum value of $(I_{max}+I_b)$, and a minimum value of $(I_{min}+I_b)$. Thus the visibility will be:

$$V'(L) = \frac{(I_{\max} + I_b) - (I_{\min} + I_b)}{(I_{\max} + I_b) + (I_{\min} + I_b)} = \frac{\delta I_S}{2(I_m + I_b)}$$

where $\delta I_s = I_{max} - I_{min}$, $I_m = (I_{max} + I_{min})/2$ and L is the position. We can also express the true fringe visibility, which does not include the background intensity, as being:

$$V(L) = \frac{\delta I_S}{2I_m}$$

Considering the two highpass coherence filtering devices, and taking the ratio between the measured (apparent) visibilities, leads to:

$$\frac{V'(L_1)}{V'(L_2)} = \frac{\delta I_{S1}}{2(I_{m1} + I_{b1})} \frac{2(I_{m2} + I_{b2})}{\delta I_{S2}},$$

If it is assumed that both interferometers are facing the same field of view ($I_{b1}=I_{b2}$ and $I_{m1}=I_{m2}$), this reduces to:

$$\frac{V'(L_1)}{V'(L_2)} = \frac{\delta I_{S1}}{\delta I_{S2}} = \frac{V(L_1)}{V(L_2)}$$

This means that taking the ratio of the measured visibilities eliminates the influence of the background in the measurement. As it is not practical to have two systems sharing the same field of view (which was assumed), one system is employed whose field of view is split into two. In this case, the upper part is like the system shown in fig. 2.5, while an additional glass delay is positioned only in the lower part. Additionally, this technique is presented in [Sut82] and [Sut89] as suitable for measurement of the coherence length of signal radiation.

2.4.2.2. Target location

By introducing a few modifications to the basic OTIM spatial coherence filtering arrangement [Sut82], it is possible to have a sensor to give the target location within the field of view. In the aperture plane a phase reticle with a linear sinusoidal thickness pattern is employed, and at the output, a four-element array (quadrant detector) is used. The variable thickness was achieved with dichromated gelatine at that time (1982), but it would be more sensible now to employ photoresist. The Fourier Transform image of the phase reticle, which rotates at the detector plane, along with the reticle geometry are shown in figure 2.8. It consists of a zero order or DC term, placed at the target position, and two rotating first order terms, the higher orders being negligible. In addition, it is one-dimensional, as the reticle has structure only in one dimension.

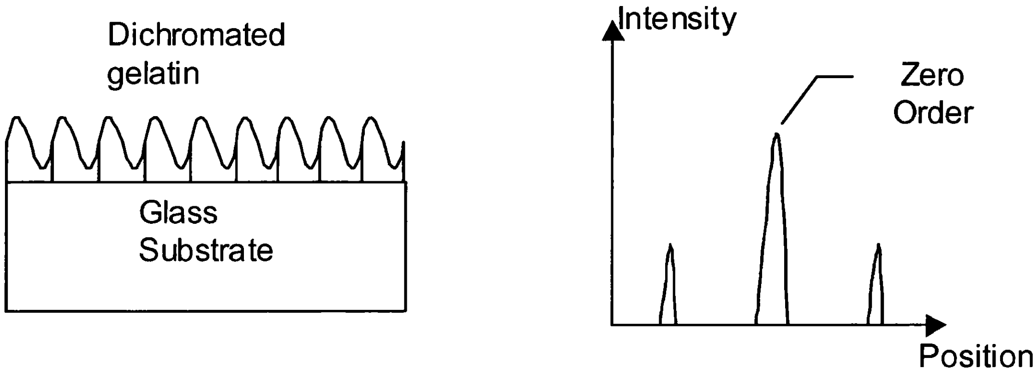


Figure 2.8 – Phase reticle geometry and its Fourier transform

Figure 2.9 shows the quadrant detector containing the Fourier transform image (FTI) of the phase reticle. By computing the time spent by the first order terms in each of the quadrants it is possible to either locate the target's position, or even derive X and Y error signals to feedback an azimuth/elevation control loop, thus keeping the target in the centre of the field of view. This is an interesting feature for both IRST and seeker systems.

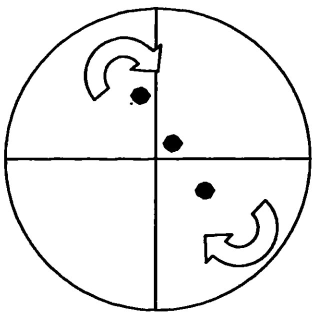


Figure 2.9 – Phase reticle FTI at the detector plane

2.4.2.3. Some concluding remarks about OTIM

A good point about the OTIM set of techniques is its broad optical band, currently very desirable in defence applications. However, this good point brings with it the problem of increased sensitivity to interfering signals in other wavelengths apart from those of interest, requiring additional processing for classification and interference rejection.

Sutton and French did not characterise the detection sensitivity of OTIM using the signal-to-clutter ratio, rather introducing a clutter rejection ratio parameter, given by [Sut82]:

$$R = \frac{\delta I_s}{\delta I_b} \qquad (2.3),$$

where δI_s and δI_b are the fringe modulation amplitudes for the signal, as shown in fig. 2.7, and for the background (residual), for an input SCR of 1. If a Gaussian target spectrum is assumed, then by calculating the Fourier transform of a Gaussian, it is possible to show that the theoretical rejection ratio becomes [Sut85]:

$$R = \frac{\delta I_S}{\delta I_b} = \exp\left[\frac{\pi^2 L^2}{2}\left(\frac{1}{L_{cb}} - \frac{1}{L_{cs}}\right)^2\right] \qquad (2.4),$$

where L is the interferometer path difference, or the coherence length filter cut-off, meaning the coherence length below which the signal is rejected; L_{cb} is the background coherence length, and L_{cs} is the coherence length of the signal. Thus if reasonable numbers are applied to eq.2.4, such as 3 μm for the background, 30 μm for the target and 10μm for L, as Sutton has done [Sut82], the theoretical rejection ratio will be approximately 193 dB. Although this is quite an impressive number, it is not feasible in practice, because the theoretical background modulation would sit below the noise floor, and this would be used to compute the denominator in eq. 2.4. Even if there was no noise at all, background modulation was found by Sutton to be larger than expected, due to aberrations in the optics, combined with the effects of extreme rays, as the FOV of the system is increased. French [Fre88] stated that a remote sensing system was implemented using OTIM, achieving a rejection ratio between 60 and 80 dB. Although no statements about minimum detectable signals were made in these works, ultimately the rejection ratio parameter can be considered as equivalent to the minimum detectable SCR, signal changed. For example, if a laser source has an input SCR of –60 dB, with a rejection ratio of 60 dB, target and background would produce the same modulation amplitudes in the output, the target being just detectable.

OTIM is a rather versatile set of techniques that can be employed in a wide range of applications where a low intensity, higher coherence (temporal or spatial) target must be recovered from a bright, less coherent background. The OTIM research carried out by French and Sutton pioneered the exploration of coherence discrimination, and did not aim [Sut82] to perform a deep investigation of particular applications, but rather to explore the whole subject in a general fashion. The work opened the way to new application-oriented approaches, such as the works of Drum [Dru90] and this thesis, among others. It is important to stress that computer technology at that time did not allow full association of OTIM techniques with digital signal processing algorithms, as is done in the approach investigated in this thesis. Even with signal processing limitations, it is likely that OTIM-based systems, once properly implemented, should be able to convert the usual background limited performance of most terrestrial EO systems into detector noise limited systems with high sensitivities.

2.4.3. Remote sensing of gases

The previous section on spectroscopy discussed the increasing interest in assessing the spatial distribution of gases, vapours and aerosols in the atmosphere, either for environmental purposes such as monitoring pollutant concentrations, or to gain a better understanding of the macroclimate phenomena at a global level. In remote sensing applications the actual concentrations of the gases cannot be directly assessed using gas cells as in the laboratory, thus the reliance on features that can be measured from sensors some distance away from the gas cloud. French [Fre84A] has described an OTIM-related set of methods of remote sensing of gases, to which both biprism and interferometer arrangements are suitable . The detection of the specimen relies on detecting changes in the temporal coherence of the input radiation, which arise from the interaction between the gas (vapour or aerosol) and an appropriate illuminating radiation. For this interaction to produce a change in the temporal coherence (or Fourier transform of the input spectrum), there must be a superposition or at least a partial overlap between a specific gas absorption line and the illuminating radiation, so that the optical bandwidth of the radiation is changed by the absorption line. The change in the transmitted or scattered spectrum is shown in figure 2.10. The reduction in optical bandwidth, causing an increase in coherence length, is represented by the horizontal arrows between the half maximum points of the spectra.

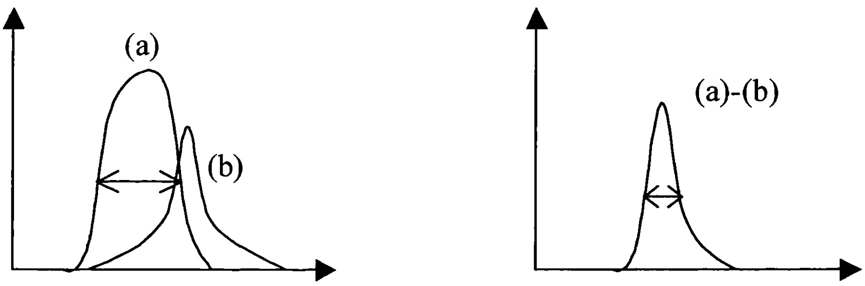


Fig. 2.10 – Optical bandwidth reduction due to the interaction of the illuminating radiation (a) with the absorption band (b)

The illuminating radiation can either be a tuneable laser, in an active mode, or natural sources, the most strong and broadband one being the sun. In the latter case, appropriate spectral filtering of the ambient illumination must be provided to limit the background energy to a narrow optical band encompassing the absorption line, imitating the narrowband artificial illumination, and obtaining a maximum change in the Fourier Transform. The cut-off frequency of the high-pass coherence filter (described in the previous sub-section), which is equivalent to the path difference bias in the interferometer, must be adjusted to a path difference where the interference fringes from the illuminating source had just been extinguished, so that even small changes in the coherence of the resultant radiation can be detected. French [Fre84A] presented the selection of such coherence cut-off as the art in the design of this kind of system, strongly affecting its detection performance.

Drum, under French's supervision, has also explored OTIM-related techniques for the same application in his doctoral thesis [Dru90]. A dedicated sensor was implemented using temporal coherence processing, and is described in [Dru87][Dru89][Dru90]. It incorporated some of the techniques from Fourier Transform Spectroscopy (FTS), described in the previous section. Unlike FTS, in Drum's research, the illuminating radiation is filtered using a sharp-edge filter, as close to rectangular bandpass as possible. Assuming its spectrum is actually rectangular, because of the Fourier Transform relationship the interferogram envelope will assume a sinc function shape. Referring to figure 2.11, rather than exploring the visibility-decaying region of the interferogram, this approach concentrates on the first minimum of this sinc envelope, a technique patented by French later [Fre93]. This interferogram-shaping process is discussed in more detail in the next chapter, where the description of the technique explored in this thesis is given.

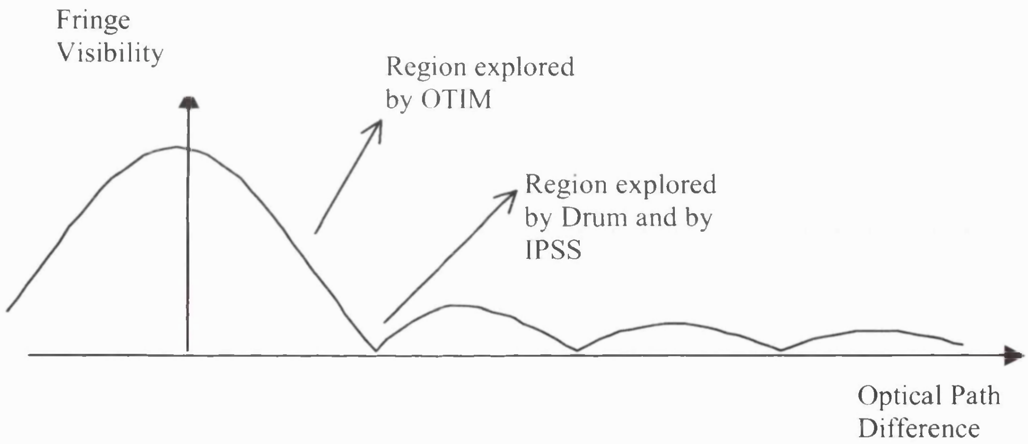


Figure 2.11 - Regions of the interferogram explored by different coherence-based approaches

The operating principle of the technique explored by Drum can be seen in figure 2.12, which we have calculated. In the upper plot, the spectrum of the illuminating radiation, assumed as flat, is presented in blue, while the spectrum of the filtered illuminating radiation (termed as background) with an absorption dip, coming from the interaction with the gas, is shown in red. In the lower plot, the same notation is used to present the corresponding interferograms. For low absorption, the change in the interferogram is proportional to the absorption line strength [Dru87]; hence, by using this method, it is possible to sense a change in the spectrum (meaning the presence of a gas) that comes into the field of view of the sensor without performing a Fourier Transform, thus without finding the spectrum. As the absorption strength is, in its turn, proportional to the gas concentration [Can99], this can be measured from the change in the amplitude of the real part of the complex degree of coherence (interferogram) at its first minimum. The region of minimum visibility was chosen because it is the most sensitive to variations in the net degree of coherence of the scene.

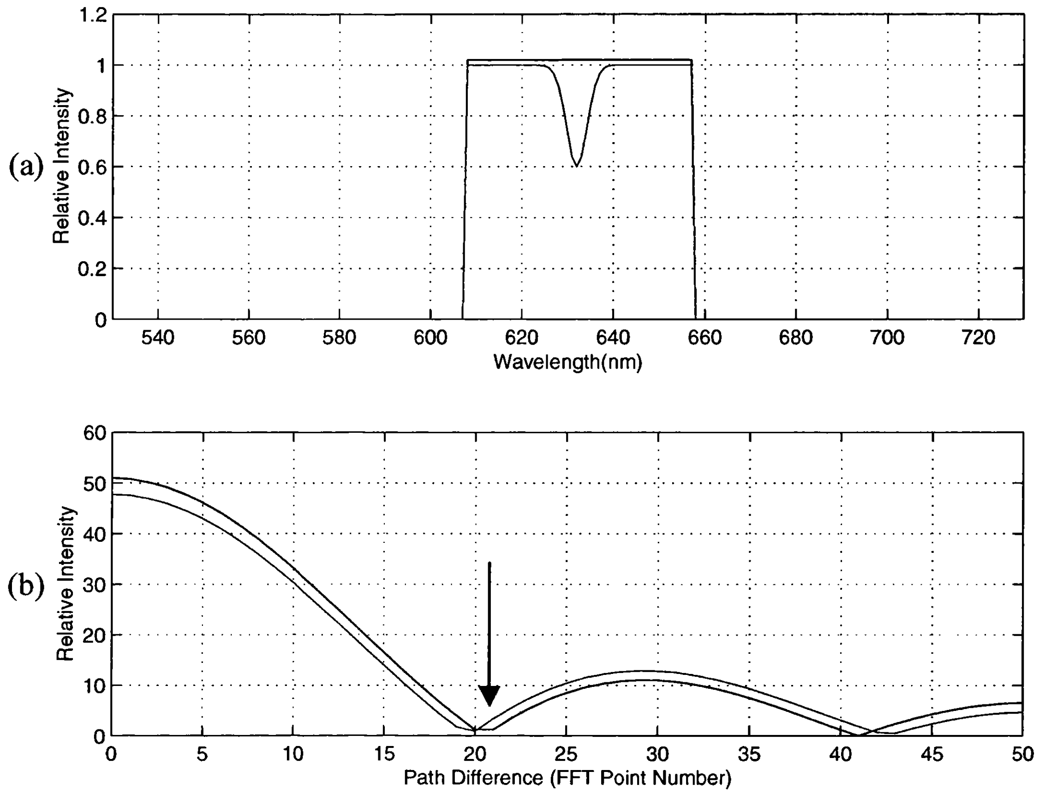


Fig. 2.12 - Change in the interferogram caused by an absorption line. Plot (a) shows the spectra, and plot (b) shows the interferogram envelopes. Blue curves represent gas absent. Red curves represent gas present. The arrow indicates the increase in coherence at the first minimum.

As pointed out earlier, it is obvious that the absorption line must fall within the optical bandwidth of the input filter, or at least some overlap must occur, so that a change in the Fourier transform can occur. Consequently, the sensor is being optimised to a specific type of gas, and, in order to detect a different specimen, another sensor would probably be required. This is another example of the use of target prior knowledge to increase detection sensitivity.

Path difference scanning is required to overcome background spectral fluctuations, pointed out by Drum [Dru90] to be the main limitation to sensitivity. The figure of merit employed in this case is the concentration-path length product, a common performance metric in gas detection systems. In [Dru89] a noise equivalent concentration path length product of 15 ppm.m for the detection of NO_2 at 489 nm was measured in the laboratory, using clear sky as the background, and a filter bandwidth of 7.5 nm. As discussed in chapter seven, this figure is considered inadequate for comparison purposes. Instead of it, a sensitivity figure taken from his thesis [Dru90] is used for comparison with IPSS in chapter seven. The use of the coherence measurement method has shown advantages in sensitivity and in faster scanning, when compared to FTS or to correlation techniques [Dru89]. Nevertheless, it has limitations coming from the need to measure the complex degree of coherence at a specific path difference, where there would be a null if no gas was present. Besides background spectral fluctuations, the uncertainties in path difference and in the coherence measurement process contribute to increase the minimum detectable concentration and, hence, degrade the sensitivity. It is claimed that the use of the real part of the interferogram offers immunity from asymmetric background spectral fluctuations, but the effect of this advantage is diminished by crosstalk from the imaginary to the real part. This theoretical immunity is a consequence of a property of the Fourier transform, according to which the even part of a function and the real part of its spectrum form a Fourier transform pair [Hsu84].

An important difference between Drum's work and FTS is that, unlike FTS, the path difference is scanned just over a small portion around the first minimum of the interferogram envelope. This brings the possibility of doing faster scanning, allowing an increase in the interferogram SNR by averaging a larger number of scans. This advantage is also incorporated in IPSS, as will be shown in chapter three.

2.4.4. Other research works on coherence detection

Another temporal coherence processing implementation, also based on OTIM, is reported by Duffy [Duf89]. It relies on the same idea of high pass coherence filtering, with two minor configuration differences: the use of the interferometer output component that is reflected back towards the source to compose the output signal, in theory doubling the system optical throughput, and the replacement of the Michelson plane mirrors by retro-reflectors made by two angle-shaped mirrors, directing the reflected flux to a detector out of the input aperture. The two outputs of the interferometer are in anti-phase [Hec98], requiring a differential amplifier to add them, as shown in figure 2.13.

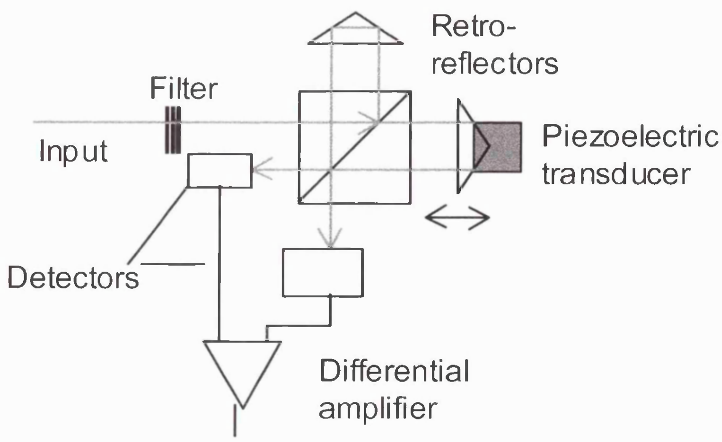


Fig. 2.13 - Temporal coherence processing employing both interferometer outputs

It should be pointed out that the use of retro-reflectors limits the system field of view, as the light paths shown by arrows in figure 2.13 are only achieved with on-axis incidence. Spectral filtering is employed in this arrangement, to reduce the amount of clutter energy causing photon noise, giving an advantage in detector signal-to-clutter ratio.

Through the combination of two sources with different coherence lengths, it is possible to measure the system rejection ratio (eq. 2.3). This was done by Duffy by setting an optical path difference in the interferometer such as to reject fringes from the less coherent source, and attenuating the most coherent one until its fringes cannot be distinguished within noise, as seen on an oscilloscope screen. Rejections of the order of 60 dB for a laser and 30 dB for a 2 nm filtered white light, using a 12 nm FWHM filter, were reported [Duf89], but this result cannot be seriously considered, as the method used for discrimination was simply the visual recognition of fringes in an oscilloscope screen. Due to the quality of the results reported, the modifications applied to the experimental arrangement, when compared to Sutton's work, appeared to be of little value. The retro-reflectors limited the useful FOV, whilst the use of two detectors doubled the shot noise, and required precise optical alignment and spectral matching between the two detected channels. These limitations are pointed out by Duffy's co-author in [Duf89] in his doctoral thesis [Hic89].

In another paper, Hickman and Duffy [Hic88] describe a proof of principle coherence-based imaging system, by replacing the photodetectors with imaging arrays, and displacing one of the interferometer retro-reflectors by half a wavelength. The pixel-by-pixel difference between the two images is taken, and, while the regions with coherent illumination present modulation, the incoherent regions are removed. The obvious limitation of this simplistic approach occurs if the image frame period is longer than the background correlation time, preventing background cancellation. The demonstration was made with two only frames, and the detection of a laser emission with an SCR of –20 dB is claimed. Chapter six presents a reproduction of this experiment, which did not obtain the claimed results.

A more recent work applied coherence filtering to build a Laser Warning Receiver [Man96]. This sensor is increasingly important for military applications, with the development of laser-guided munitions. The interferometric device is now a Fabry-Perot etalon [Hec98], made from an electro-optic crystal. The crystal is an electro-optic material, varying its refractive index with the applied voltage. The etalon is made by the deposition of layers of different refractive indices, creating a Distributed Bragg Reflection (DBR) effect. The successive reflections create a transmissivity response in the form of the Airy Function, as shown in figure 2.14. If the sought after laser signal linewidth is smaller than the Airy peak FWHM, it is possible to adjust the etalon thickness such that the laser line falls within one of the cavity modes, or Airy function peaks (and that is one limitation of the approach). Fine and fast tuning of the etalon will move those peaks, thus changing the cavity transmissivity for the laser radiation, and modulating the laser output intensity. On the other hand, incoherent signals will not be modulated because of their broad bandwidth, thus being rejected.

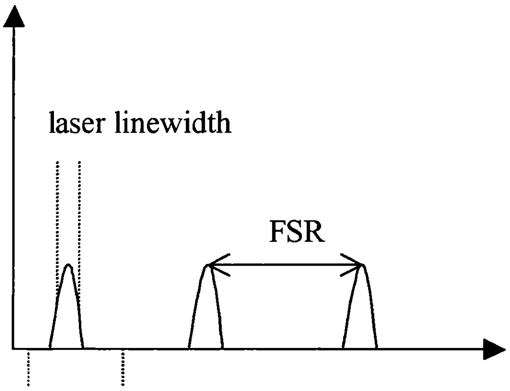


Fig. 2.14 - Laser warning receiver principle. FSR is the free spectral range

The system employs the Fabry-Perot resonator as its principal component. Its output is fed to a photodetector, whose electrical output is bandpass filtered at a frequency corresponding to the modulating frequency of the etalon. This then goes to a threshold circuit and to an alarm in case detection is declared. In the detection of CW lasers a periodic structure of InSe was modulated at 100 Hz; for a nanosecond-pulsed laser, the modulation period has to be shorter than the pulse duration, and InSe has a slow response time. Hence, the design changed to a $LiNiO_3$ crystal placed inside a microwave cavity driven at 1 GHz, not using electrodes and avoiding parasitic capacitance problems. The etalon resonator was drilled with 3mm holes, so that microwave electric field changed the etalon refractive index, thus giving the modulation effect required. The main problem with this design was microwave oscillator leakage into the optical signal, which was reduced by coupling photodetector and electro-optic cell through an optical fibre. The minimum detectable laser signal was observed in a microwave spectrum analyser, being limited by the interference signal from the microwave source. Using this detection criterion, the minimum detected laser power is 10 mW, and the radiation from a 100 W light bulb placed in the input aperture is not detected. This does not mean that the minimum detectable SCR is –40 dB, because, although not explicitly mentioned, it can be inferred from the paper that no controlled light bulb power measurement was performed. If this is true, then in the experiment only a fraction of the 100 W total power reached the system input aperture, and the SCR would be a lot higher. Besides the limitations previously presented about visually recognising a modulation in the output signal, another problem would be found in designing a filter to detect partially coherent radiation, which inevitably would produce much smaller modulation indexes.

This example shows that a Michelson Interferometer is not necessarily the best design solution for every particular problem, the choice being governed by factors like characteristics of the signal to be detected, amount of optical path length scan required, weight, volume and others. It also shows that electro-optic or opto-acoustic cells allow much faster modulation schemes to be employed.

2.4.5. Optical Coherence Tomography (OCT)

In the previous sub-sections, coherence-based detection methods were presented that take advantage of the higher degree of coherence of a target, when compared to the background. In the past ten years there has been an intensive development of interferometric techniques that exploit the low coherence of a light source, in order to detect or image reflected radiation coming from sub-surface layers of transparent media. The set of techniques named Optical Coherence Tomography (OCT) [Sch99] [Rei99] originated from research on white light interferometry and optical coherence domain reflectometry (OCDR), the latter recently being used in test equipment for location of discontinuities in optical fibres. In these instruments, coherence has replaced both time and frequency domain reflectometry with resolutions some three orders of magnitude smaller, and twice the dynamic range [Agi01]. Although the physical principles and arrangements are similar, OCDR and OCT differ in that the first is concerned with one-dimensional ranging, while the second is employed in imaging sub-surface layers of live tissue. In its simplest mode, OCT uses a broadband light (low coherence) source to illuminate an interferometer, whose reference arm is a mirror as in FTS, and the sample to be tested sits in the second arm, replacing the second mirror. In the FTS interferometer the width of the self-coherence function [Bor99] or interferogram is inversely proportional to the optical bandwidth of the radiation in the input aperture. If a broadband source is employed (white light), a very narrow pulse in the coherence domain is produced at zero path difference (see chapter four for an experimental plot of this pulse). Through adjustment of the position of the reference mirror, OCT and OCDR use this narrow pulse to control the depth within the sample where the reflected coherence pulse will come from. The spatial resolution of the technique is, thus, limited by the width of this pulse, the shorter the coherence length, the better.

OCT limitations arise from the interaction of the beam radiated into the tissue with its scatterers and absorbers. Not very much is known about this interaction, and about the origin of the speckle usually seen in the images [Sch99]. To date, penetration depths reach 2 cm in the eye, and 1-2 mm under the skin, using near-infrared wavelengths, and dynamic ranges reach 90 dB [Sch99]. The use of free space rather than fibre-based interferometers is favoured because they do not suffer from dispersion and changes in polarisation that reduce the amplitude of the coherence peak.

In a recent magazine review article [Rei99], Reiss states that the key areas for OCT improvement are the light source, image acquisition rates and data management software. The light sources usually employed are edge-emitting LEDs, superluminescent diodes and dedicated lasers, with coherence lengths ranging from 7 to 60 μm. Research in this area aims to obtain lower coherence lengths, with higher power, at reasonable costs. Image acquisition usually occurs in real time, but not video rates, ranging from 4 to 20 frames-per-second (fps). IPSS, investigated in this thesis, can offer an advantage in the light source issue, as low coherence is not critical, and in the other two areas mentioned, as smaller scans are required, generating less data.

Novel imaging modes are being incorporated into OCT for better sensitivity. Polarisation-sensitive OCT [Sch99][Hit99] uses the fact that muscle and tendon tissues can exhibit birefringence when aligned in layers, and has been used to measure depolarisation by multiple scattering [Sch99]. Another interesting development by Hotate [Hot00] uses a stepwise frequency modulation of a laser source, to synthesise a self-coherence function made of a comb-like sequence of pulses. By changing the laser modulation frequency, the path difference at which the pulses appear can be tuned, so that the lowest order pulse lies in the depth of interest, within the imaged object. Hence, path difference mechanical scanning is not required. One of the limitations of the technique comes from the stepwise frequency change, which generates coherence pulses with a sinc format (Dirac impulses in the limit), whose sidelobes decrease the depth resolution, and cause image blurring. In a recent magazine article [Hot01], Hotate applies this synthetic coherence function method to the measurement of strain in an optical fibre, by investigating its Brillouin gain spectrum distribution, with a spatial resolution 100 times better than conventional, time domain techniques.

2.5. Conclusion

This chapter has presented a review of techniques that are related to IPSS in two different senses. Sections 2.1 and 2.2. presented techniques applying to the detection of dim targets in heavy clutter, which is one of the strengths of IPSS, to be demonstrated later in this thesis. Sections 2.3 and 2.4 presented FTS and OCT respectively, techniques that, although not primarily concerned with detection, display architectures very similar to the one of IPSS.

For clear comparison, table 2.1 presents detection figures of merit claimed by the authors of the papers reviewed, in order of appearance in the text. Because of the different applications, different types of figures are given. Although the majority of the examples are used in chapter seven for comparison, some of them were not included there as, due to the lack of information about the experimental conditions, it was not possible to obtain figures of merit common to these approaches and to IPSS. The table presents the source of the result in the "reference" column, the figure of merit used by the authors, as they have named it, the numerical results reported, and a brief summary of the conditions in which the results were obtained.

Reference	Figure-of-Merit	Result	Conditions
[Tar00]	Clutter rejection	16.5 dB	Simulated data
[Bra96]	Minimum SNR	5 dB	Simulated data
[Wei95]	Minimum SCR	8.2 dB	Simulated data, single frame
[Wei95]	Minimum SCR	6 dB	Simulated data, 4 frames, target maximum speed 1 pixel/frame
[Tar99]	Minimum SNR	-6.6 dB	Simulated target, real background, P_{fa}=0.017, 20 frames, 20 sec delay
[Dia01]	Minimum SCR	0 dB	Laboratory data, 16 frames averaged
[Mon00]	SNR gain	52 dB	Field data
[Sin98]	SCR gain	15.5 dB	Not clear
[Sch96]	SCR gain	8 dB	Two bands
[Sch96]	SCR gain	17 dB	Three bands
[Sch96]	SCR gain	21 dB	Four bands
[Eis97]	Clutter rejection	30 dB	Temporal FTS, simulated data
[Che96]	Minimum SNR	-3 dB	Simulated data, laser wavelength measurement
[Thé01]	Minimum Concentration-path length	258 ppm.m	Real data, remote sensed at 105 m
[Nel00]	NETD	0.3 Kelvin	Imaging FTIR, 11 μm, real data, 16 scans
[Nel00]	NETD	1.0 Kelvin	Imaging FTIR, 11 μm, real data, 2 scans
[Nel00]	NESR	3.2 μflicks	Imaging FTIR, 11 μm, real data, 1 cm^{-1} resolution, 1 scan
[Fre88]	Rejection Ratio	60 dB	Conditions not clear; experiment not described
[Dru89]	Minimum Concentration-path length	15 ppm.m	Visible band, in the laboratory, using the sky as the background, 7.5 nm filter bandwidth

[Duf89]	Rejection Ratio	60 dB	Experiment in the laboratory, laser target, 12 nm filter, detection criterion is the visualisation of modulation
[Duf89]	Rejection Ratio	30 dB	Experiment in the laboratory, 2 nm filtered white light target, 12 nm filter, detection criterion is the visualisation of modulation
[Hic88]	Minimum SCR	- 20 dB	Experiment in the laboratory, subtraction of two image frames

Table 2.1 - Detection figures of various approaches

Analysing the data in table 2.1, the minimum SCR detected using intensity-based techniques is –6.6 dB, although it takes 20 seconds to do so. This example shows the most important drawback of such techniques: large amounts of data are generated, requiring very potent computing resources, many times yielding unacceptable time delays. While some researchers keep trying to improve detection sensitivities by building complex algorithms, some others have recognised that only the use of information from other domains can improve performance [McC96]. Indeed, the inclusion of the spectral domain in detection systems has improved detection performance. Clutter rejection ratios from 8 to 52 dB were obtained with such systems, this figure being proportional to algorithm complexity, hence to the amount of data generated and processed. Another way of exploring the spectrum is FTS, actually measuring it with very high resolution. Initially developed for the laboratory, FTS techniques and equipment are now being adapted to face the real-time operation requirements in the remote sensing and defence fields. The great challenge for such an operation is the large amount of data FTS generates, posing problems on data processing and management.

There is also a requirement for sensor broadband operation. This allows the use of a single sensor to detect a wide range of targets, either gases in a remote sensing system, or threats in a defence application. OTIM techniques pioneered the use of coherence information, embedded in the light wave that reaches the sensor, achieving 60-80 dB clutter rejection with an optical bandwidth limited by the optical components. It is clearly different from the techniques discussed previously in that this performance is achievable with modest signal processing, not requiring Fourier transformation. FTS and OCT are broadband as well, but unlike OTIM do require a lot of computing power.

In the review a few examples of the use of prior knowledge about the target to be detected to simplify the detection process are given. The coherence measurement technique applied by Drum [Dru90] to the remote sensing of gases associated the high sensitivity of OTIM with the benefits of narrowing the optical band of the instrument. The optical filtering not only increased the SCR, but also created a feature in the interferogram, that avoided the need for a long interferogram scan. The investigation of a single point in the interferogram was required. Path difference scanning was employed just to overcome background spectral fluctuations. On the other hand, the approach requires accurate calibration of path difference, to make sure the right point of the interferogram is being measured, and of the coherence measurement process as well, depending on the intensity of the real component of a filtered interferogram. Whenever an intensity is measured, one has to consider detector dynamic range, amplitude noise and integration times.

When compared to coherence measurement, IPSS relies on measuring times rather than intensities. It is believed, and this is demonstrated in chapter seven, that IPSS compares favourably with most spatial, spatial-temporal and multispectral techniques, being more sensitive, and easier to implement; with FTS and OCT, for being simpler and faster; and with coherence measurement, for not having its calibration problems. The penalty for such an improvement is the need for prior knowledge about the target to be detected (as happens with coherence measurement). In any application where such knowledge is available, it will be shown that IPSS can offer a simpler, faster and more sensitive solution.

Chapter 3

Theory

This chapter presents, in the form of a brief review, the theoretical background required for an understanding of the developments described in this thesis. It is not a comprehensive review of the fields covered, but rather a collection of definitions and concepts that are often used throughout the text, in an attempt to make the thesis a stand-alone document. As the technique studied in this thesis explores the fields of optics and signal processing, the sections of the chapter cover both areas, and are organised as follows: the first section concerns optics, reviewing concepts of coherence and interference, as well as the influence of polarisation on both phenomena. The second section is about signal processing, and describes the tools used in the recovery of the instantaneous attributes of a signal, such as the Hilbert transform, the analytic signal and the complex envelope. The final section describes Interferogram Phase Step Shift (IPSS), the main subject of this thesis, using contributions from both fields. Almost all concepts in this section already existed at the start of the author's research, with the exception of the use of a frequency spike as an event marker, instead of a phase step. The novel algorithm used to extract this spike from the interferogram is not presented here, but in chapter seven. The advanced theory concepts, used to model and predict the performance of the detection system, are shown in chapter eight.

3.1. Coherence and interference

3.1.1. Introduction

The concept of coherence is tied to that of correlation [Gar92], and is used in many different fields that deal with fluctuating quantities, such as optics, image processing, communications, radar, atmospheric science and astronomy. In a milestone review paper published in 1965 [Man65], Mandel and Wolf postulate: "*In its broadest sense, optical-coherence theory is concerned with the statistical description of the fluctuations, and optical coherence phenomena may be said to be manifestations of correlations between them*". The definition reinforces the fact that statistics is an essential tool in coherence theory, and that coherence is a form of correlation. Coherence phenomena may be described in two ways: using classical wave theory, where the fluctuating fields are treated as waves, or using quantum mechanics, where they are treated as quantised entities. Both approaches have been shown to be equivalent [Man65]. In this thesis, the use of coherence theory is limited by the following assumptions:

- Use of the classical, wave approach;
- Study of electromagnetic waves of light;
- Use of second-order coherence, that is, the study of correlation between two points in space and time. This formalism allows an unified treatment for coherence and polarisation phenomena [Man65];
- Study mainly of self-coherence phenomena, that is, between two points of the same wave field, with less emphasis on mutual coherence, which assesses the correlation between two or more fields;
- The fields studied are stationary and ergodic (defined in [Pap91]), allowing the use of time averages to replace ensemble averages. Furthermore, they are described as complex, fluctuating functions of time, and are assumed to be *quasi-monochromatic*, or with an optical bandwidth much smaller than its central wavelength. This latter assumption is justified by the use of narrowband interference filters and light sources in the experiments described in chapters four to six.

The coherence between fields can be classified into three types [Gar92]: temporal, where the correlation is measured between the values of the field in one point in space, at two instants in time; spatial, where this measurement is performed at two different points in space, but at the same time; and spectral, which applies to the computation of mutual coherence between frequency-shifted versions of a field, and is not studied in this thesis. The relationship of the first two types with polarisation can be visualised by considering two points in an electromagnetic wave, which is a transverse field, propagating in the +z direction of an orthogonal co-ordinate system. The correlation between the fields at the two points in the direction of propagation is the temporal coherence, and depends on the wavefront propagation time between the points. The correlation in the plane perpendicular to the direction of propagation is the spatial coherence, and depends on the distance between the points. The cubic region in space formed by two spatial and one temporal dimensions, where a certain amount of coherence exists is called *coherence volume* [Man65]. Furthermore, the correlation between the x and y components of the field in either point is its polarisation [Bor99]. Therefore, polarisation, spatial and temporal coherence are part of the total correlation between the fields. This relationship is discussed later in this section.

3.1.2. Interference

The first investigations of coherence came with experiments performed in the nineteenth century to study interference phenomena [Man65]. Among several such experiments, the most representative (to the subject of this thesis) were conducted by Young (1801) and Michelson (1885), and are widely described in the literature [Hec98][Bor99]. Young investigated interference phenomena that depended on the spatial coherence of the light source, while Michelson invented the interferometer that bares his name, and studied interference exploring the source's temporal coherence. The Michelson interferometer was described in chapter two, and was used in all the experiments in this thesis.

The interference phenomenon was described by Hecht [Hec98] as *"...an interaction of two or more light waves yielding a resultant irradiance which deviates from the sum of the component irradiances"*. When two monochromatic, point sources separated by a large distance (compared with their wavelengths) generate linearly polarised waves that overlap at the same point in space, in accordance with the Principle of Superposition, their electric field vectors combine, so that the net field is given by the sum of the two vectors. These fields can be represented, using classical electromagnetic theory notation, by:

$$\mathbf{E}_{1,2}(\mathbf{r},t) = \mathbf{E}_{o1,2}\cos(\mathbf{k}_{1,2}.\mathbf{r} - \omega t + \phi_{1,2}),$$

where $\mathbf{r}$ represents the spatial co-ordinates and t the time of measurement, $\mathbf{k}$ is the wavevector, whose modulus equals $2\pi/\lambda$ and orientation indicates the direction of propagation of the wave; ω the angular frequency assumed to be the same in both waves, and ϕ is the phase of each field, relative to a common reference. $\mathbf{E_1}$ and $\mathbf{E_2}$ are vectors with different orientations. As the light intensity, expressed through the calculation of the Poynting vector, is given by [Hec98]:

$$I = \varepsilon v < E^2 > ,$$

where ε is the electric permittivity of the medium, v is the phase velocity of the wave in it and the brackets represent a time average. By adding the fields, squaring and employing a trigonometric relation [Hec98], the total intensity I_T is given by:

$$I_T = \frac{E_{01}^2}{2} + \frac{E_{02}^2}{2} + \mathbf{E}_{01}.\mathbf{E}_{02}.\cos\delta \qquad (3.1),$$

where ε and ν have been dropped out, and δ is the net phase difference between the waves, given by:

$$\delta = (\mathbf{k}_1.\mathbf{r} + \phi_1 - \mathbf{k}_2.\mathbf{r} - \phi_2) \qquad (3.2)$$

In eq. 3.1, the first two terms are the intensity of the individual light beams, and the last term is called the interference term, which accounts for the interaction between the fields of the two waves and is the term of interest. In this equation, it is also possible to see that the amplitude of the interference term varies in accordance with the scalar product of their amplitude vectors, or with the cosine of the angle between their orientations. From this fact, one may conclude that the difference in polarisation direction between two fields reduces the amplitude of the interference term, even eliminating it if the fields are orthogonal. The phase difference in eq. 3.2 also determines the amplitude of the interference term. The term $\mathbf{k}_1.\mathbf{r}$-$\mathbf{k}_2.\mathbf{r}$ is caused by the path difference that might exist between the interfering fields, while ϕ_1-ϕ_2 is due to their original phases. In order to produce observable interference, the latter term has to be constant, requiring the two phases to be tied. This is achieved through interferometers, where a single light source is used and the beam is split and recombined to interfere with a delayed version of itself. The optical path difference [Hec98], then, can be expressed as $\tau = \frac{\lambda}{2\pi}.\delta$. A plot of intensity vs. path difference (for example fig. 3.4 in section 3.3) is known as an *interferogram* [Hec98], a term widely used in this thesis. In the case of monochromatic waves, the interferogram is an endless sinusoid, with a period corresponding to the wavelength. This is quite intuitive as, when the two waves are exactly in frequency and phase, their amplitudes add totally, and they are said to cause constructive interference; when they have a half-wave phase difference, total destructive interference occurs. The variations in intensity, when seen on a screen are called bright fringes for constructive interference, and dark fringes for destructive interference.

3.1.3. Partial coherence

In the previous sub-section, some assumptions were made in order to present the interference phenomenon. The beams were assumed to be generated by point sources, so that the effects of spatial coherence could be neglected [Bor99]. In a classical Young's experiment, extended sources can be thought of as an infinite set of elementary point sources, each one generating a fringe profile. As a result, the net fringe pattern, as seen on a screen, will be composed by the overlapping of spatially shifted elementary interferograms. The resulting intensity pattern would then depend on the correlation between the elementary fields emitted by the point sources, which decreases as the distance between them increases. The value of this distance, for which a pre-determined reduction in the amount of correlation occurs, is called the *spatial coherence length*.

Another assumption was that the light source was monochromatic. Polychromatic sources can be decomposed into wavelength components that, without non-linearities, will interfere separately as monochromatic waves. The interference patterns will be superimposed at zero path difference, because there is no phase delay, but as this path difference increases the various individual profiles will show different periods of the above mentioned sinusoidal interferogram, and the range of path differences over which the addition of these elementary interferograms add coherently will be larger if the range of wavelengths present in the optical source radiation is reduced. At large path differences, the elementary interferograms just average out to the first two terms in eq. 3.1, as the interference term becomes negligible. This range of path differences over which fringes occur is the *temporal coherence length*, and, as discussed above, it is inversely proportional to the optical bandwidth of the source. The time the wave takes to travel the temporal coherence length is the *coherence time (T_C)*.

A different interpretation of coherence length can be given by thinking of optical radiation as a set of wavetrains [Hec98], whose emission is associated with the physical mechanism behind the generation of the radiation. For a laser, the presence of a stable resonant cavity forms longer wavetrains, characteristic of the laser's spectral purity; for a white light source, wavetrains of various and random frequencies, phases and duration are emitted, and their phases can only add coherently within the length of a wavetrain. The coherence length can be thought of as the average length of these wavetrains, and the duration as the *coherence time*. Hence, a laser has a longer coherence length, and a narrower optical bandwidth; white light has a very short coherence length, because it is broadband. This intuitive inverse relation between coherence length and optical bandwidth can be expressed as [Bor99]:

$$L_c = \frac{\lambda_0^{\,2}}{\Delta\lambda} \tag{3.3}$$

where L_c is the coherence length, λ_0 is the mean wavelength, and $\Delta\lambda$ is the optical bandwidth. A more detailed discussion of coherence length is presented in the next sub-section.

3.1.4. Measures of coherence

As described previously, the nature of polychromatic light waves is non-deterministic, meaning that there is no way to describe their behaviour using analytical expressions. Instead, they are a set of various wavetrains of deterministic signals, whose distribution in amplitude, phase and frequency can only be appropriately described with the aid of statistic variables. Although the instantaneous fluctuations of the field cannot be measured with present detectors, by means of interference experiments one can measure their correlation properties and, indirectly, retrieve information about the field. There is a strong analogy between the cross-correlation and the auto-correlation functions, used in statistics to characterise random variables, and the mutual coherence and self-coherence functions, used in the study of optical fields. Using this analogy, if one replaces the random variables with two optical signals, $E_1(t)$ and $E_2(t)$, then the degree of mutual coherence between the two signals is the magnitude of their temporal correlation coefficient [Gar92], which in optical jargon is known as the *complex degree of coherence* of the optical signals:

$$\gamma_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)}.\sqrt{\Gamma_{22}(0)}} \tag{3.4}$$

where Γ_{12} is the (non-normalised) mutual coherence function of the signals, defined as $\Gamma_{12}(\tau) = < E_1(t+\tau).E_2^*(t) >$ [Bor99], and Γ_{11}, Γ_{22} are the self-coherence functions, employing the same expression. For $\tau=0$, one has $\Gamma_{11}=I_1$ and $\Gamma_{22}=I_2$, the intensities of the radiation from both sources. $\gamma_{12}(\tau)$ is a complex function, with amplitude and phase. The phase is intensively exploited in our Interferogram Phase Step Shift (IPSS) technique, as will be shown in section 3.3. If one takes the real part of $\gamma_{12}(\tau)$, it is possible to obtain an equivalent form of eq. 3.1:

$$I = I_1 + I_2 + 2\sqrt{I_1.I_2}.\mathrm{Re}(\gamma_{12}(\tau)) \tag{3.5}$$

This is known as the *General Interference Law for Partially Coherent Light.* The modulus of $\gamma_{12}(\tau)$ ranges from zero to one; at longer path differences, white light has a very low degree of coherence, and a gas laser usually has a very high degree of coherence. Most radiation sources, in nature or man-made, range in between these two limits. By applying eq. 3.5 to calculate the maximum and minimum intensities that would be possible to obtain from the interference between two optical beams, it is possible to find:

$I_{\max} = I_1 + I_2 + 2\sqrt{I_1.I_2}.|\gamma_{12}(\tau)|$, when the phase of $\gamma_{12}(\tau)$ is zero; and

$I_{\min} = I_1 + I_2 - 2\sqrt{I_1.I_2}.|\gamma_{12}(\tau)|$, when it is π. In the situation where $I_1=I_2$, found in an interferometer using a beamsplitter with an even splitting ratio, the visibility of the fringes, defined as $V(\tau) = {(I_{\max} - I_{\min})}/{(I_{\max} + I_{\min})}$, is equal to the modulus of the complex degree of coherence. The visibility can then be interpreted as a function or signal that modulates the fringe profile, giving it a particular envelope. Hence, for a perfectly coherent source, this envelope would never decay with increased path differences; on the other hand, the visibility curve of a white light source decays very rapidly. Example interferograms demonstrating these properties are shown in chapter four.

It is worthwhile now to re-assess the concept of coherence time using the theory already presented. Essentially, the coherence length is the maximum path difference at which a light source can produce interference fringes, but the visibility decay is gradual, and no sudden cut-off exists. Furthermore, the relationship between the coherence length and the optical bandwidth, intuitively addressed as a reciprocal law in eq. 3.3, is not deterministic, as from one it is not possible to fully determine the other. Neither the spectral density nor the self-coherence function of a non-deterministic field has a fixed width, such as in a top-hat function. Instead, the spectra can be Gaussian, Lorentzian, or a combination of the two. In statistical theory, the Weiner-Khintchine [Hec98] theorem states that the Fourier transform of the auto-correlation function is the power spectral density function [Pap91]. In the study of optical fields, the same relationship holds, or *"the Fourier transform of the self-coherence function of a radiation field is its power spectrum"* [Hec98].

Due to the non-deterministic nature of the coherence length and the optical bandwidth, Born and Wolf [Bor99] have defined the coherence length as the normalised root-mean-square (RMS) width of the square modulus of the complex degree of (temporal) coherence:

$$L_c^2 \equiv \frac{\int_{-\infty}^{\infty} \tau^2 . |\gamma(\tau)|^2}{\int_{-\infty}^{\infty} |\gamma(\tau)|^2} \qquad (3.6),$$

where for simplicity γ is presented without the indices. The definition of coherence length is definitely not agreed upon by researchers [Man65], and, as well as the ones given by eqs. 3.3 and 3.6, Mandel proposed a third one [Man65][Goo85]:

$$L_C \equiv \int_{-\infty}^{+\infty} |\gamma(\tau)|^2 \, d\tau \qquad (3.7)$$

Comparing the three definitions, the one given in eq. 3.3 is very limited, in the sense that it requires knowledge of the optical bandwidth, whose definition is also disputable. The most popular (and simple) criterion used in optics is the Full Width Half Maximum (FWHM), measured between the points with an amplitude of half of the peak amplitude. Along with the definition given by eq. 3.6, Born and Wolf present the concept of effective spectral width, as the normalised RMS width of the square of the spectral density:

$$(\Delta f)^2 = \frac{\int_0^\infty (f-\overline{f})^2 . G^2(f).df}{\int_0^\infty G^2(f).df}, \quad \text{where} \quad \overline{f} = \frac{\int_0^\infty f.G^2(f).df}{\int_0^\infty G^2(f).df}$$ is the mean frequency. This definition is better, because it is applicable to any power spectrum. When comparing Wolf's (eq. 3.6) and Mandel's (eq. 3.7) definitions, care must be taken, as for light sources with spectral structure, the formulae might lead to fairly different results, as will be shown in chapter four.

More rigorously, the relationship between the coherence length and the effective spectral width follows the same form as the one found in Heisenberg's uncertainty principle [Bor99]:

$$T_c.\Delta f \approx \frac{1}{4\pi} \qquad (3.8),$$

where T_C is the coherence time previously defined and the $\approx$ sign means " the same order or magnitude". The exact value of the term the right hand side of eq. 3.8 depends on the light source's spectral profile. In general, without any assumptions, the only thing that can be said is that this value is more than $1/4\pi$ or 0.08 [Man65]. If a Lorentzian profile is considered, the factor equals 0.318; for a Gaussian, 0.664; and for a rectangular (top hat) spectrum, it equals unity [Goo85]. The top hat approximation is used in eq. 3.3, demonstrating how poor the approximation can be when studying spectral profiles of real light sources.

3.1.5. Temporal coherence, spatial coherence and polarisation

In section 3.1.3, the effects of spatial and temporal coherence were discussed separately. In reality, these two manifestations of coherence exist together, influencing the shape of the visibility functions (and interferograms). If the light sources used to record the interferograms display *cross-spectral purity*, then the effects of temporal and spatial coherence can be evaluated separately [Goo85]. Basically, a light source has cross-spectral purity if the power spectrum of its radiation does not vary across the area of the beam.

Polarisation affects the interference phenomenon by reducing its efficiency, when compared to linear, coincident polarisations. Coherent light is always polarised, as the amplitude and phase of the electric field do not vary randomly. It this type of radiation, as shown in eq. 3.1, the degree of mutual coherence between two fields (proportional to the fringe visibility) is reduced according to the cosine of the angle between their polarisation orientations, if linear. If the light is partially polarised, only the polarised portion will impair the mutual coherence. The theory of partial polarisation is similar to that of partial coherence [Man65]. The coherency matrix formalism [Bor99], not detailed here, can be applied to quasi-monochromatic light, and allows the calculation of the degree of polarisation of the wave, from measurements performed with phase compensators and polarisers, much in the same way as the degree of coherence can be assessed from intensity measurements performed with interferometers.

The separation between temporal and spatial coherence, and the effect of polarisation, then, allow one to write the *net complex degree of coherence visibility*, or the fringe visibility considering all three effects, assuming beams of the same intensity, in the following form adapted from [Man65]:

$$\gamma_N(\tau,d,\theta)=\gamma_S(d).\gamma_T(\tau).\cos(\theta_P) \qquad (3.9),$$

where γ is the complex degree coherence, with the subscript N for net degree, S for spatial, and T for temporal, τ is the path difference, d is the distance, on the surface of the light source, between the two points whose coherence is measured, and θ_P is the angle between the direction of polarisation of the linearly polarised fields at the overlap region. Eq. 3.9 intends to show that the fringe visibility depends on the temporal and spatial coherence of the light source, as well as of any polarisation difference between the fields. The effects of polarisation on the experiments presented in the thesis are discussed in chapter eight.

In summary for this section, the definitions of coherence length are not universally agreed, and their use is governed by the spectral profile under study. There is a reciprocal relationship between coherence length and optical bandwidth, but the exact factor of (inverse) proportionality also depends on the spectral profile. The Fourier transform of the self-coherence function is the power spectrum. Its normalisation gives the complex degree of coherence, whose modulus equals the visibility function, which, in its turn, is proportional to the interferogram envelope. Temporal and spatial coherence, as well as differences in polarisation states, affect this envelope and, under certain conditions, can be analysed separately.

3.2. Calculation of the instantaneous attributes of a signal

A key component of the Interferogram Phase Step Shift (IPSS) technique is the recovery of the instantaneous phase (and frequency) of an interferogram, which can be considered to be a signal in the time domain. This task is performed through the phase demodulation of the interferogram signal, and relies on a set of mathematical calculations used in the description of narrowband signals. A narrowband signal is one that has a spectrum concentrated around a relatively high carrier frequency [McD98], having a negligible amount or no energy at zero frequency. Consider a narrowband signal, recorded, for example, using a digital oscilloscope. This signal would consist of a sequence of real samples. The calculation of its instantaneous amplitude and frequency would require the generation of another sequence, with complex samples. The instantaneous amplitude and phase signals would then be formed by the modulus and phases of the complex sequence, respectively.

The following derivation of the instantaneous phase of a narrowband signal was adapted from Proakis and Manolakis [Pro96]. Consider the spectrum of a narrowband signal *x(t)* with two bands, centred at +/- f_c, the carrier frequency. Its Fourier transform represents its spectrum *X(f)*. The signal obtained by removing the negative frequency contents of *X(f)* is called the *analytic signal* [Pro96]. In the frequency domain, it is obtained by filtering out the negative part of the power spectrum, and doubling the result to keep the same power. The frequency response of this filter equals then twice the unit step function. In the time domain, the analytic signal, represented by $x_+(t)$, is the result of the convolution between *x(t)* and the inverse Fourier transform of twice the unit step:

$$x_+(t) = [\delta(t) + {}^{j}\!/_{\pi t}] * x(t) = x(t) + {}^{j}\!/_{\pi t} .* x(t) = x(t) + j.\hat{x}(t) \qquad (3.10),$$

where $\hat{x}(t) = \frac{1}{\pi t} * x(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau}.d\tau$ is the *Hilbert transform* of *x(t)* [Pro96]. The filter which implements the transform has a frequency response with unity amplitude, and a phase of -π/2 for positive frequencies and +π/2 for negative frequencies. Hence, in the frequency domain, the Hilbert transformer is a 90° phase shifter.

Eq. 3.10 shows that the analytic signal is a complex signal whose real part is the input real signal, and the imaginary part is its Hilbert transform. As it is still a band-limited signal, in order to obtain the instantaneous attributes of the input signal, a down-conversion is required. In the frequency domain, one simply calculates the baseband version of the analytic signal $X_+(f)$, $X_L(f)=X_+(f+f_c)$. In the time domain, the analytic signal has to be multiplied by a complex exponential centred on a frequency f_c to obtain its lowpass version, called the *complex envelope* $x_L(t)$. Representing this signal as $x_L(t)=a(t)e^{j\theta(t)}$, the instantaneous amplitude of the original signal is *a(t),* and its instantaneous phase is $\theta(t)$.

3.3. Interferogram phase step shift (IPSS) technique description

3.3.1. Introduction

This section presents a description of the technique under investigation, starting with the location of the IPSS technique within the approaches described in chapter two. Intensity-based systems detect and integrate the energy radiated by target and background within the sensor optical bandwidth, usually limited by the detector spectral response. Dual band systems use, instead, two different spectral regions selected by optical filters, and can achieve higher sensitivities. The number of spectral bands can be increased to tens or hundreds, reaching, in the limit, the very high spectral resolutions achievable with Fourier transform spectrometers (FTS), at the cost of computational complexity, required to manipulate large amounts of data, and long processing times [Che96]. In FTS, the spectrum is calculated from measurements in the coherence domain, namely, from interferograms, via Fourier transformation.

A way to reduce the complexity of the computational algorithms involved in the detection process is to extract the required information in the coherence domain, without performing the Fourier transformation. This is done in techniques like Optical Coherence Tomography (OCT) [Sch99] and Low Coherence Interferometry [Hit99], allowing imaging with smaller computational costs. The Interferogram Phase Step Shift (IPSS) technique, which is the main subject of this thesis, allows further reduction in the amount of data to be processed, by limiting the path difference scan range required to take the measurements to a small region around an event marker intentionally created in the interferogram. At the highest level, IPSS can be described as a technique that looks for incremental increases in the coherence degree of a scene, being capable of sensing a very small change in a spectrum, in terms of total power, without measuring it directly.

3.3.2. Aspects of IPSS that already existed before the research described in this thesis

To start describing the technique under investigation, the main concepts of FTS are briefly reviewed. In FTS the coherence profile is scanned through a range of path differences as long as possible, as the spectral resolution achieved from the Fourier transform is the reciprocal of the length of the scan [Cha79], ranging from a few microns for white light sources, to centimetres, required to appropriately display a laser line. In real time applications, such as the detection of a missile or in biomedical imaging, high update rates are often required. The long scan, usual in FTS, followed by a Fourier transform would take a long time, which might render the technique not applicable to such applications. IPSS can perform a faster scan of the interferogram, by continuously sampling a special localised feature it might have, and looking for changes in its location to declare a detection. A narrowband interference filter, placed at the detection system's input, creates this feature.

Recalling the definition of coherence length, and its relationship to optical bandwidth given in section 3.1, it is noticeable that a source such as a laser will produce fringes over a very long range of path differences, up to a few centimetres in the case of gas lasers for use in the laboratory. This interferogram envelope (or self-coherence function) will look like a horizontal straight line. At the other extreme, white light (or background radiation) interferes with itself for a very short range of path differences over about 1 to 3 microns, depending on the optical bandwidth of the optics used to measure the interference. The introduction of a filter with a relatively narrow optical bandwidth increases the coherence as it reduces the bandwidth (see eq. 3.3), so that the interferogram dies down at longer path differences.

If an input optical filter with a steep roll-off characteristic is used, closely resembling a rectangular spectral response, and assuming that the input radiation spectrum is flat within the filter bandwidth, the profile will assume a sinc (*sin(x)/x*) function shape, according to the Fourier transform relationship between power spectrum and self-coherence function. The sinc function nulls are chosen to be the special features already mentioned. In fig. 3.1, a rectangular filter spectrum is inverse Fourier transformed, giving the modulus amplitude of the self-coherence function in the upper plot and the phase of the enclosed carrier (fringes) in the middle, and its instantaneous frequency, obtained by differentiation of the phase, in the lower plot. The phase goes through π (180°) steps at the positions of the minima, which are chosen to be the event markers used for detection. Such phase steps (or jumps), shown in the middle plot of fig. 3.1, as well as their usefulness as event markers, were first introduced by French [Fre93], and are widely explored in this thesis. The use of the phase step is claimed by French to be much more accurate than the amplitude minimum, as in this region the fringe visibility is very low, and additive noise may severely affect the event marker. More recently, Lobachev and others [Lob01] described the use of this feature as a reference signal in Low Coherence Interferometry. The phase jump occurs because the sine function in the numerator of the sinc-shaped self-coherence function changes sign whenever the argument of the sinewave is a multiple of pi radians. In a scheme where this self-coherence function is measured as an interferogram, using a photodetector, there is a jump in the phase of a signal that is measured as a photodetector power reading. The occurrence of this phase feature in power is not surprising, if one recalls that the detector power reading is proportional to the coherence function, which, in its turn, displays the phase feature. Another interesting remark is that the phase jump is measurable if either of the intensities of the interfering beams, represented by the first two terms in eq.3.5, are removed before the measurement, or not, as the phase jump is contained in the interference term.

The linear decay observed in the phase profile is due to the fact that the rectangular spectrum employed for Fourier transformation is not centred at zero frequency (or wavelength), a property of the Fourier transform [Hsu84]. A physical interpretation of the phase of the interferogram is the relative position of the fringes, if spatially distributed on a screen. The phase step would, then, be an alteration of the period at which the fringes usually appear. This alteration in period is shown as real images in chapter six. The plot in fig.3.1 c) is discussed in the next sub-section.

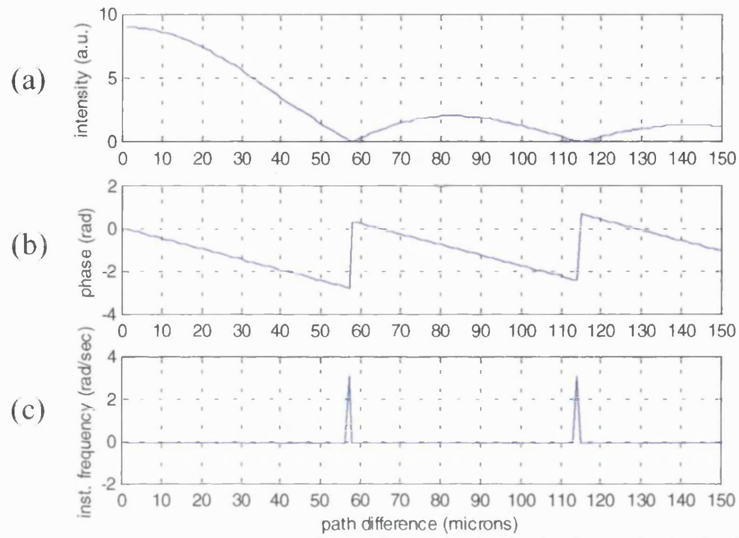


Figure 3.1 - Self-coherence function of a flat spectrum transmitted through a rectangular filter response, showing modulus amplitude (a), phase (b) and instantaneous frequency (c)

3.3.3. Novel aspects of the IPSS technique

It was shown in chapter two that most processing in previous coherence-based approaches was done before the conversion of the radiation into an electrical signal. In IPSS we have used more intense post-detector signal processing, in addition to the optical pre-detector processing. Along with the phase step invented by French [Fre93], we have investigated the use of a novel event marker derived from it. The differentiation of the phase steps produces frequency "spikes", which, we believe, are more easily detectable than the phase steps. Firstly, the differentiation removes the slope of the phase function, which prevents the use of an amplitude threshold. Secondly, as we are interested in sudden phase changes, the differentiation acts as a highpass filter, increasing the relative size of these changes, and discarding low frequency variations. The width and height of the spikes are determined by the rate of change of the phase. In theory, this rate is infinite, and the spikes are Dirac impulses. However, in real signals, the phase transition is not actually a step, having a finite rise time, and the sharpness of the frequency spikes is limited by factors like the electrical bandwidth of the detector and electronics, analogue-to-digital conversion sampling rate and noise. Chapter seven presents a detailed study of the detectability of the spikes. The frequency spike is chosen, then, as the event marker, instead of the phase step, facilitating the implementation of the computer algorithm to extract it from the real, noise-degraded interferograms, also presented in chapter seven.

The recording of the exact position of the frequency spike corresponding to the first minimum of the envelope yields a reference value, in path difference units. The first minimum is chosen as the event marker because of its higher fringe visibility when compared to the others, allowing a better SNR. When a coherent source moves within the field of view of the detection system, its narrow spectral feature changes the incident spectrum, thus changing the interferogram. It can be either an emission, as in the case of lasers and low-pressure lamps, or an absorption, when the illuminating radiation propagates through a region containing molecules that absorb some specific wavelengths, according to their rotational and vibrational resonant modes [Can99]. As will be demonstrated in chapter eight, the insertion of this narrow spectral feature in the field of view changes the position of the minima in the interferogram, as shown in figure 3.2. The reference position of the interferogram minimum is shown in blue. If this feature is an absorption, shown in red, the displacement is negative, or towards lower path differences. If it is an emission (shown in black), than the displacement is positive.

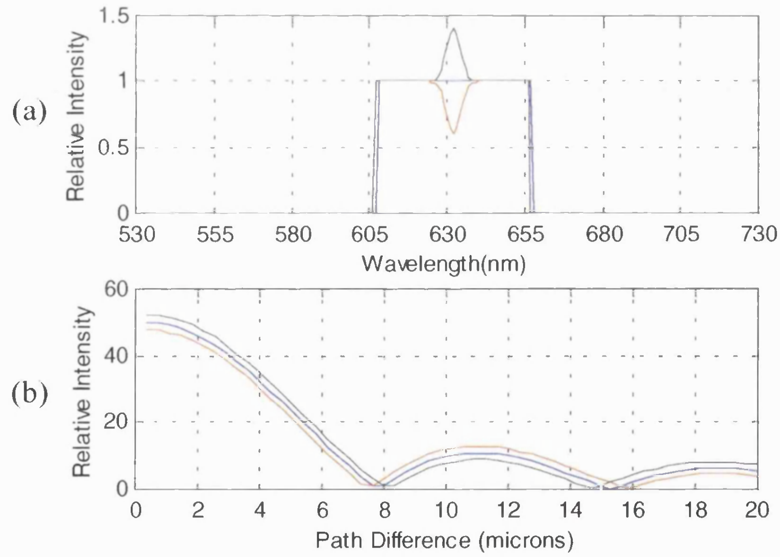


Figure 3.2 - Displacement of the interferogram minimum. Plot (a) shows the spectra, and plot (b) shows the self-coherence functions. The reference position (no target present) is seen in blue; an emission target is in black and an absorption target is in red.

By tracking the position of the event marker previously mentioned, a detection can be declared whenever this position changes by an amount larger than a specified threshold. The threshold level will, then, determine the probabilities of detection and false alarm of the system, the signal being the displacement of the frequency spike due to the presence of the coherent target, and the noise being all other sources of displacement, such as changes in the background spectrum, and other sources of noise. It is important to point out the main difference between IPSS and the Coherence Measurement method, reviewed in chapter two. In the latter, the change in *amplitude* of the real component of the self-coherence function at the original position of its first minimum is measured; in the former, a change in *position* of this minimum is detected. The fact that the actual degree of coherence is not actually being measured brings advantages over that technique, reducing the constraints on path difference uncertainties reported by Drum [Dru90]. Another advantage of IPSS (over dispersive spectrometric techniques), is the incorporation of the Fellgett and Jacquinot FTS advantages, described in chapter two. Even compared with FTS, it has the advantage of employing faster scanning, reducing the effects of noise by integration of various scans, providing higher update rates, and generating less data.

A typical implementation of an IPSS detection system would require an elementary set of components, shown in fig. 3.3. After being collected by the input optics, the radiation is filtered by a bandpass interference filter with the steep roll-off characteristic, already described. The interferometer acts as a correlator and measures the self-coherence function (or interferogram) corresponding to this filtered radiation, which has the event marker formed in it by the interference filter. The interferogram radiation is then focused on the surface of a photodetector, which converts the radiation into an electrical signal. Electronic circuitry conditions the signal for digitisation, and a computer algorithm aims to extract, from the digitised interferogram, the position of the event marker, as well as any displacements it might suffer. As shown in fig. 3.3, the approach comprises optical, electronic and digital components, and one of its strengths actually comes from its interdisciplinary layout.

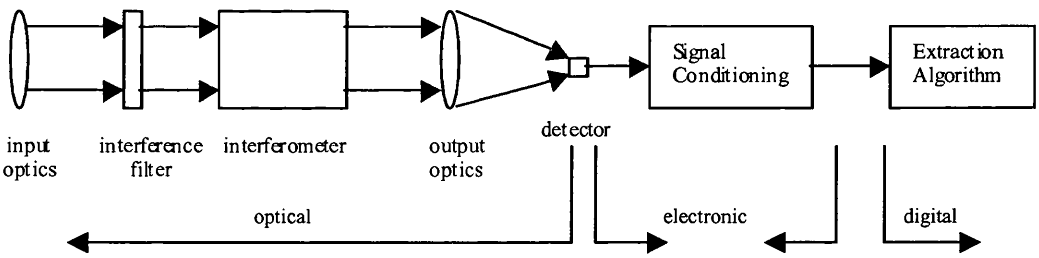


Figure 3.3 - IPSS main components

It is important to give some more detail about IPSS's main components, namely, the interference filter, the interferometer, and the extraction algorithm. Concerning the interference filter, and considering a remote sensing application as a thought exercise, the background radiation reaching the detection system is incoherent, with a spectral shape corresponding to the combination of self-irradiated and sun-reflected components, each of those governed by Planck's equation and by emissivity or reflectivity variations due to the physical characteristics of the different materials in the scene. This broadband background spectrum is affected by absorption by atmospheric components, mainly CO_2 and H_2O, the first one between 4.1 and 4.3 μm and the second in bands about 2.7 and 6.3 μm. By carefully selecting an appropriate central wavelength and bandwidth of an optical filter it is possible to sample only the region of interest, where target features are likely to be present, and also rejecting background clutter out of the region. Nevertheless, the most important role of the filter is to shape the interferogram, to create minima at specific positions, as described previously. In order to produce sidelobes that give the minimum to be examined, the filter must have a spectral response as close to rectangular as possible. The sharp frequency response edges are met with interference filters, which can be implemented using multilayer cavities.

Real interference filters will not actually have a rectangular spectral response, but very steep roll-off characteristics can be achieved with Fabry-Perot cavities. For example, one of the filters used in chapter five has a roll-off of 620 dB/decade (in a frequency scale). Its fractional bandwidth is 1.7%. One feature of these filters is their asymmetrical spectral response in relation to the central mean wavelength, which is due to changes in the reflectivity of the cavity mirrors with wavelength. As a property of the Fourier Transform [Hsu84], this will cause the interferogram envelope to have no zeroes, reducing the slope of the phase change at the minima, making the frequency spikes smaller, which reduces the detectability of the frequency spikes. Therefore, a symmetric response is always desirable.

Other important parameters of the filter are its optical bandwidth and central wavelength. The latter has to be matched, as closely as possible, with the mean wavelength of the target to be detected. There is an optimal optical bandwidth that gives IPSS the maximum responsivity to a given target. If the filter is too narrow, it gives the background a coherence length, which is comparable to that of the target, reducing the coherence contrast and, consequently, the amount of its change with the presence of a target. If the filter is too wide, in comparison with the target, besides passing much more background power than necessary, its self-coherence function might be so narrow that its frequency spikes would be very close to zero path difference, becoming undetectable. Chapter eight studies these two issues in a deeper level.

An optical interferometer has to recombine amplitudes or wavefronts previously separated, but coming from the same source. Those portions of the original radiation will then interfere, if the difference between the optical paths travelled is less than the coherence length. As mentioned in section 3.1, this interference signal (interferogram) is proportional to the self-coherence, or the auto-correlation function, of the input radiation. Hence, from this point of view, the interferometer is a correlator. In the detection system studied in this thesis, a Michelson interferometer is employed, operating with collimated beams. This has the advantage of allowing the wavefronts to interfere in the whole area of the optical components, increasing the optical throughput, and reducing the effects of extra path differences created by extreme rays, linked to poor collimation. If the light is not well collimated, extreme rays leave the collimating optics at an angle with the optical axis, travelling longer than the principal rays, and generating undesirable path differences [Cha79]. Further details on the Michelson interferometer were given in chapter two.

An interesting feature of IPSS is that it can be combined with post-detector processing, producing a very high-resolution system, and incorporating the advantages of both optical and digital signal processing. On detection, the optical signal arising from the interferometer is converted to an electrical one. Unlike conventional systems, this detection is keeping information about coherence and spectrum of the source, imposed on it by the correlation operation [Sut82]. Due to the linear scan of the optical path difference, and provided that the source has enough coherence, the radiation intensity at the output of the interferometer will vary sinusoidally, as shown in eq. 3.1. If this fringe profile fills the whole output aperture, a detector placed at the back focus of a focusing lens will convert this intensity variation into an electric signal, consisting of the sinusoidal fringes modulated by the self-coherence function of the radiation, as shown in figure 3.4. This signal is amplified by an instrumentation low noise pre-amplifier and lowpass filtered, to give a reasonable analogue signal-to-noise ratio. The amplified signal is then fed into an analogue-to-digital converter (A/D), where it is quantised and coded into a sequence of bits. The sampling frequency must obey the Nyquist theorem [Ope89] to avoid aliasing of the sampled signal, and, additionally, be high enough to pick-up the phase step, whose frequency is higher than that of the carrier coming from fringe scanning. A calculation of this carrier electrical frequency is included in chapter four. The digital signal is demodulated in phase, where its instantaneous phase and frequency are derived. In fact, any means of recovering the instantaneous frequency of the interferogram are suitable, such as a Phase-Lock loop (PLL), for example.

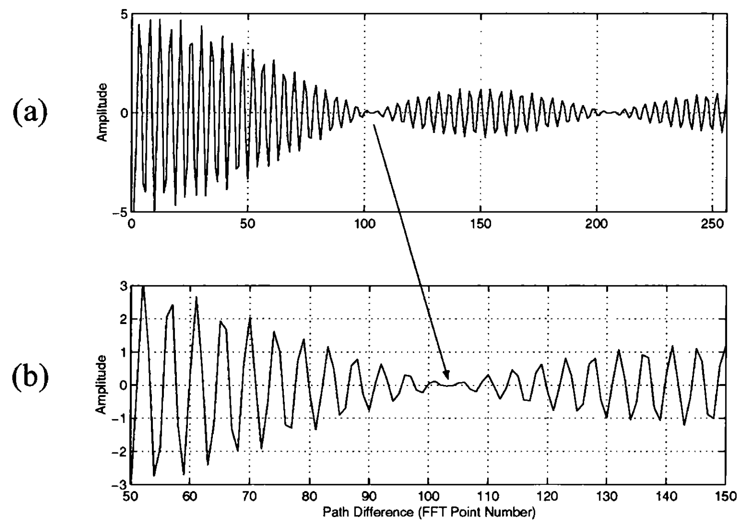


Figure 3.4 – A long section of the IPSS interferogram (a) and its minimum visibility region (b)

After calculating the interferogram's instantaneous frequency, whose theoretical profile was depicted in fig. 3.1 c), the system then finds the path difference at which this instantaneous frequency reaches a maximum, with the target absent, which is taken as a reference, as well as with the target present, and the difference (shift) is calculated. This offers the advantage of not needing to find the zero path difference, neither requires path length calibration using a laser, which usually are important issues in FTS [Cha79]. What needs to be assessed is the change of the path difference at which the phase step occurs. This means that, rather than an absolute path difference measurement, just a relative measurement is required. The computer algorithm used to perform phase demodulation and to calculate the phase step shift is detailed in chapter seven.

In conclusion, the IPSS technique incorporates the sensitivity advantages of spectroscopic approaches, when compared to intensity-based approaches. Additionally, it has the FTS advantages, if compared to dispersive techniques. When compared to FTS, it is faster and produces smaller amounts of data, with reduced computational effort. One disadvantage arises from the use of narrowband filtering, which reduces the available optical bandwidth considerably. The description of the theoretical background to IPSS proceeds in chapter eight, where the advanced theory is presented.

Chapter 4

Coherence Length Measurements

This chapter represents a slight deviation from the main subject of the thesis, optical detection, and explores the measurements of coherence lengths of partially coherent sources. This deviation is justified by the need to characterise the optical sources used in the experiments in terms of coherence length, to assess the relationship between this length and the sensitivity of the Interferogram Phase Step Shift (IPSS) approach. A secondary objective was to gain knowledge of the characteristics and limitations of the interferometric system employed in the IPSS experiments, presented in chapters five and six. As this is the first experimental chapter in the thesis, it starts with a brief description of the interferometric system alignment procedure, then describes the method chosen for the measurements (direct, interferometric), and displays full interferograms obtained from different sources. In the next section, a computer algorithm developed to calculate the coherence lengths is explained. A case study is then presented, that involved the measurements of coherence lengths of RCLED (resonant cavity light emitting diode) devices kindly furnished by the Centre for Electronic Materials and Devices, Imperial College, and their variation with different parameters. A conclusion section summarises the achievements and completes the justification of the inclusion of this work in the thesis.

4.1. Interferogram recording

The initial experimental arrangement consisted simply of some light sources, described later, collimating optics and a Michelson interferometer (Coherent Universal Interferometer, catalogue number 25-9093). The alignment of an optical system based on an interferometer needs to be started with manual path difference adjustment, to allow for fringes to be visualised on a screen, before moving one of the interferometer's mirrors automatically. Initially, an expanded He-Ne laser beam was used as a light source, as due to its longer coherence length, fringes could be observed in a wide range of path differences. Its mounting was adjusted to ensure that its beam had a constant height. In a Michelson interferometer, the two mirrors need to be exactly perpendicular in order to obtain circular fringes [Hec98]. As shown in chapter three, the instrument only displays fringes if the optical path difference is less than the coherence length of the incoming radiation, hence, the zero path difference (ZPD) position must be found. A coarse adjustment was made by changing the path difference in order to enlarge the first order (central) fringe. The fine adjustment was done with a less coherent source, an LED, and finally with a tungsten halogen bulb, for which the interferogram is very narrow in path difference and locates the ZPD accurately. The path difference was changed until coloured fringes could be seen within a path difference range of a few microns. These fringes represent a white light interferogram, where different colours of visible light interfere constructively at different path differences. Fig. 4.1 shows the fringes from white light as displayed on a screen placed 30 cm from the instrument's exit aperture.

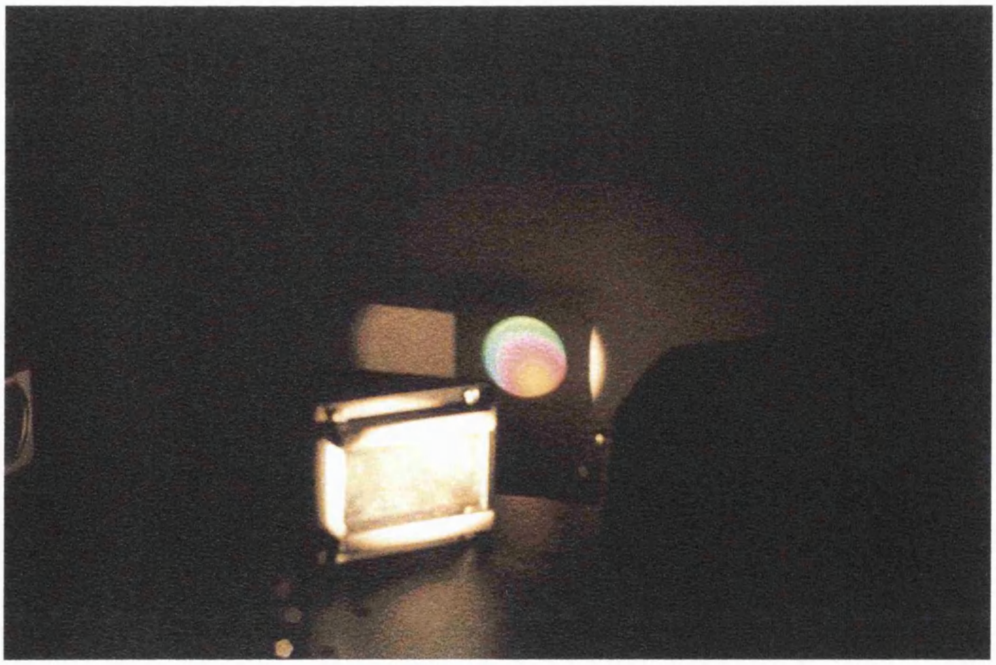


Figure 4.1 - White light fringes seen in a screen placed at the exit of the interferometer

After these procedures, the system was ready to produce interference fringes from various sources. In order to measure the fringe patterns or interferograms, the detector was placed at the centre of the pattern, whose intensity was varied by changing the path difference using a stepper motor driven manually through a knob, and recording the detector readings. The interferometer's tilting mirror was adjusted to give fringes wider than the detector active diameter (1 cm), but due to instrument's imperfection it was not possible to achieve fringes centred with relation to the exit aperture, as seen in figure 4.1. By varying the path difference manually, elementary interferograms from three sources with different degrees of coherence were measured. Light from a 20 W tungsten-halogen bulb, with a blackbody temperature of 3220 K, filtered by an interference filter with a central wavelength of 647 nm and an optical bandwidth of 12.2 nm, gave the longest interferogram; an ultrabright LED with a central wavelength of 650 nm and a bandwidth of 18 nm produced a set of fringes shorter than the previous one; and unfiltered light from the tungsten halogen bulb gave an even shorter interferogram. These measurements provided confidence to the automation of the interferogram recording process.

In this thesis, the recording of interferograms required means of varying the path difference between the two beams in the interferometer, of converting the optical fringe signal into an electrical one, and of digitising this electrical signal at the output of the photodetector for signal processing. The path difference was varied temporally, as in Fourier transform spectroscopy (FTS) [Cha79], by moving one of the interferometer's mirrors, using either a piezoelectric transducer, or a micropositioner. The detector employed was a silicon photodetector (Newport 818-SL), operated in the photovoltaic mode, whose output was connected to a Tektronix 2430A oscilloscope via an audio amplifier. The data stored in the internal memory of the oscilloscope were then transferred to a computer via an IEEE-488 standard interface. Further details about the experimental arrangements are given in chapter 5, which describes the IPSS experiments.

In the first automated arrangement, the path difference was changed using a piezoelectric transducer, connected to a piezoelectric driver (Melles Griot 17PCS001). Assuming linearity, the conversion factor between sample number, defined as the sequential number of a sample in the interferogram data block, and path difference, is calculated using:

$$\frac{PathDifference}{Sample} = \frac{PathDifference}{DrivingVoltage} \cdot \frac{DrivingVoltage}{Time} \cdot \frac{Time}{Division} \cdot \frac{Division}{Sample} \qquad (4.1)$$

Its terms represent the processes within scanning and acquisition. Starting from the end of the equation, the oscilloscope memory buffer has 1024 samples, which fill 20 horizontal screen divisions, giving a factor of 0.0195 division/sample; the time/division factor corresponds to the reciprocal of the oscilloscope time base in divisions/sec. The ratio (driving voltage)/time is determined by the characteristics of the driving waveform, in this case triangular, and as the equation assumes linearity, it represents the scan rate in volts/sec. The drive voltage range was set at the maximum (75 volts), and the rising time was 330 ms, for the 1.5 Hz scanning frequency employed. The first term (path difference/driving voltage) is a characteristic of the piezoelectric transducer. By changing the driver voltage and counting the number of wavelengths passing on a screen (one wavelength per fringe), this factor was found to be 0.488 μm/V. It is worth noticing that the product of the first two terms represent twice the moving mirror speed, and the product of the latter two represent the sampling interval, reciprocal of the sampling rate. Placing these products in eq. 4-1 yields:

$$\frac{PathDifference}{Sample} = 0.732.\frac{R}{T}, \tag{4.2}$$

where R is the oscilloscope time base, and T is the triangular waveform rise time.

Eq. 4.1 can also be used to calculate the mirror speed and the expected electrical frequency of the fringes in Hz, as measured by the oscilloscope. The mirror speed, assumed to be constant during the rise time of the driving waveform, is given by:

$$V = 0.244.\frac{75}{T} = \frac{18.3}{T}, \tag{4.3}$$

where the first factor is the mirror displacement in μm/V (half of path difference change), and the second is the maximum driving voltage. The electrical frequency of the fringes is then the path difference change, times the number of wave cycles per unit length, or the wavenumber. As the former is twice the mirror speed, one has:

$$f = 2.V.\bar{\kappa} \tag{4.4}$$

where f is the fringe frequency in Hz, and $\bar{\kappa}$ is the mean wavenumber of the filtered radiation, reciprocal of wavelength.

Fig. 4.2 presents interferograms from a tungsten halogen lamp (a), the same ultrabright LED used previously (b), and the tungsten halogen lamp light filtered by an interference filter with a mean wavelength of 632.8 nm and an optical bandwidth of 11 nm (c). The horizontal axis of the plots were normalised by using equation 4.1 and placing the axis origin at the maximum intensity fringe (assumed to correspond to ZPD). The interferograms were recorded using a 1.5 Hz triangular waveform, and 100 ms/div oscilloscope time base, with the exception of the white light, which used 10 ms/div in order to expand the narrow coherence profile. The plots are in agreement with the reciprocal relationship between optical bandwidth and degree of coherence, shown in chapter three. Due to its low degree of coherence, the tungsten halogen bulb interferogram (a) decays very rapidly. An intermediate behaviour is seen in the LED interferogram (b), while the interference filter (c) produced fringes up to the 36 μm limit of path difference that can be scanned with this transducer, established by its maximum driving voltage. The slight increase in amplitude at the end of the scan was due to non-linearity in the transducer. The corresponding coherence lengths of the optical sources are calculated in the next section to be 1.16 μm for the bulb ("white light"), 5.44 μm for the LED, and 12.59 μm for light filtered by the interference filter.

(a)
(b)
(c)

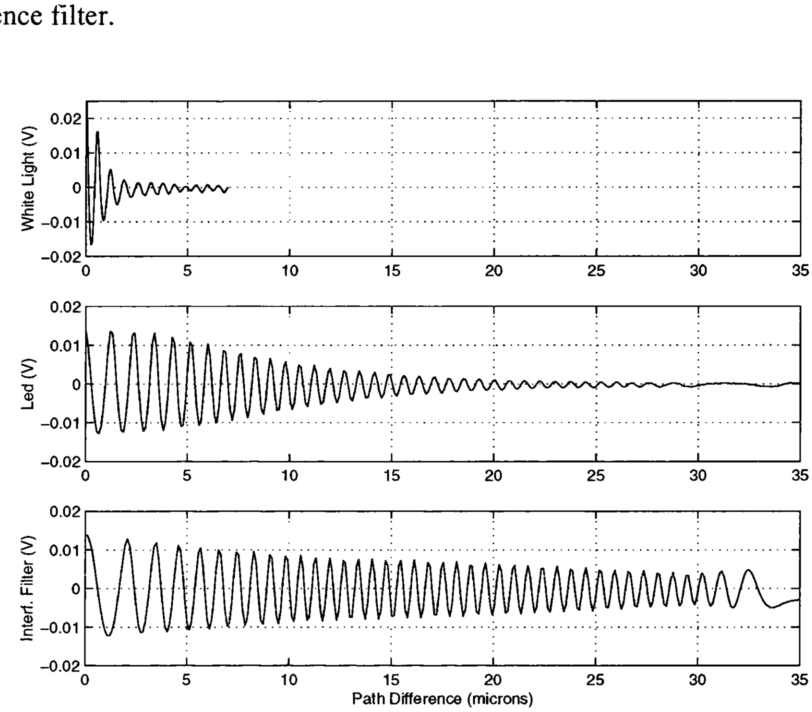


Figure 4.2 - Interferograms from (a) a tungsten halogen bulb,(b) an InGaAlP LED, and (c) an interference filter. The vertical scale in the plots is the photodetector reading.

In order to assess the properties of the interferograms generated by interference filters, a larger two-sided interferogram containing the minimum visibility regions on both sides, from the 11 nm wide, 632.8 nm-centred filter illuminated by the 20 W tungsten halogen bulb was recorded. It was made from five partially overlapped sections, joined by matching the fringes of neighbouring sections that had the same amplitude, and neglecting the overlapping fringes. The resulting interferogram is presented in figure 4.3. The mirror scanning frequency was 2 Hz, and the time base was 50 ms, yielding a sampling rate of 1.024 kHz. The minimum fringe visibility at positive path difference is at 36 μm, in good agreement with the result of 36.4 μm, obtained using eq. 3.3, but the negative minimum is at –31 μm. If the position of the positive minimum is taken as the half width of the curve, represented by the average between the two minima, it is at 33.5 μm, within 8% of the theoretical value. An asymmetry enlarging the left portion of the interferogram is observed. These imperfections are due to the limitations of the simplified splicing procedure, and to non-linearities in the path difference scanning by the piezoelectric transducer. The plot is useful to observe the behaviour of the self-coherence function close to its minimum, the region explored in Interferogram Phase Step Shift (IPSS).

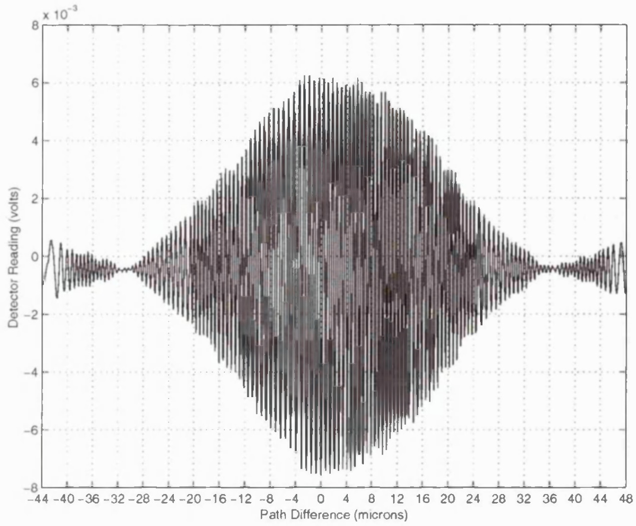


Figure 4.3 - Double-sided interferogram for a 632.8 nm centred, 11 nm FWHM interference filter

The transmission spectrum of the filter used for these measurements, as supplied by the manufacturer, is shown in fig. 4.4; its Fourier transform (amplitude and phase), linear spectrum phase assumed, is presented in fig. 4.5 (notice the abrupt phase transition used as a marker in IPSS). Assuming that the tungsten halogen lamp spectrum is flat within the filter optical bandwidth, the Fourier transform of the filter transmission spectrum is proportional to the interferogram envelope, in both phase and amplitude, as shown in chapter three. The horizontal axis of fig. 4.5 was calculated as follows: the horizontal axis resulting from the FFT is converted to a baseband frequency, depending on the interferogram sampling frequency; the baseband frequency is converted to wavenumbers using eq. 4.4, and the frequency in wavenumbers is converted to wavelengths using their reciprocal relationship. The first minimum occurs at a path difference of 34.3 μm, quite close to the one given by the splicing procedure (33.5 μm). The difference is possibly due to temperature differences between calibration and measurement situations, and to the assumption that the amplitude of the complex degree of coherence equals the fringe visibility curve, which only holds if the beamsplitter has an even splitting ratio. The beamsplitter used in these experiments had a splitting ratio of 46-54%.

Due to the path difference range limitation of the piezoelectric transducer employed, the interferometer was modified to have one of its mirrors mounted on a DC motorised translation stage, with a resolution of 0.1 μm, connected to a digital controller (Klinger MC4). This controller was connected to a computer via an IEEE 488 interface, where a program in Labview language can control the movement of the stage in any pre-programmed fashion. The path difference conversion from sample numbers to μm is, in this case, given by:

$$\frac{PathDifference}{Sample} = \frac{PathDifference}{Step} \cdot \frac{Step}{Time} \cdot \frac{Time}{Division} \cdot \frac{Division}{Sample} \qquad (4.5)$$

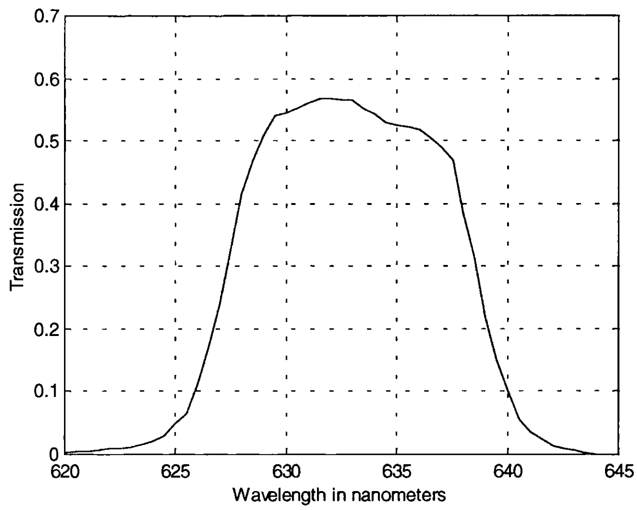


Figure 4.4 – Interference filter transmission characteristic. Central wavelength is 632.8 nm, FWHM is 11nm.

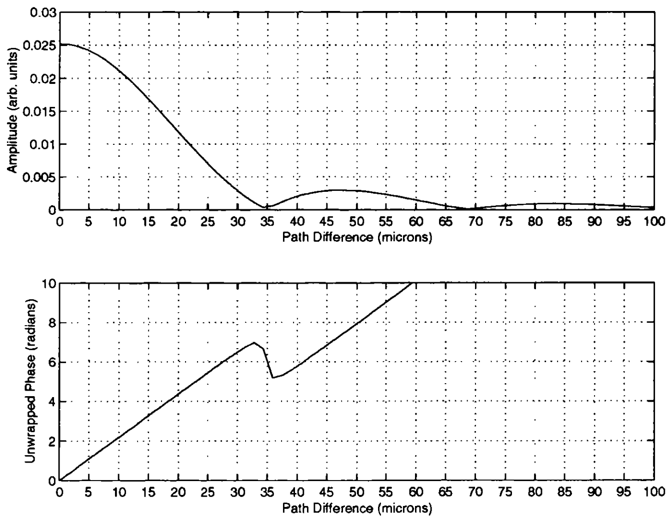


Figure 4.5 - Corresponding interferogram amplitude and phase

The first term on the right side of the eq. 4.5 is 0.2 μm/step, as each step in the stepper motor that moves the translation stage corresponds to 0.1 μm in mirror displacement, or 0.2 μm in path difference. The second term depends on the number of steps per second set by the computer program, up to a maximum of 1500 steps/sec, and sets the mirror speed.

Fig. 4.6 presents an interferogram recorded using the translation stage arrangement. The stage speed was 500 steps/sec, corresponding to 100 μm/sec path difference change. The translation stage embodiment has the advantage of performing much longer scans, in this case 400 μm, but at the expense of higher noise than in the piezoelectric transducer case. By experimentation, it was found out that this noise was caused by the stepwise displacement of the translation stage, and to electromagnetic induction from the stepper motor in the electronics, and could not be removed by filtering, as its frequency was very close to the fringe carrier signal. Although the fringe carrier frequency was calculated to be 181.82 Hz, using eq. 4.4, spectral analysis of the signal shown in fig. 4.6 displayed a spectral peak around 50 Hz. This peak was the effect of aliasing of the carrier, whose frequency (181.82 Hz) was 53.82 Hz higher than half the sampling frequency (128 Hz). This problem happened with the interferograms recorded for coherence length measurements, presented in the next section, that had a sampling frequency of 256 Hz. Due to hardware limitations of the digital oscilloscope used in the recordings, in order to acquire the whole interferogram signal, at the frequency it was generated, a relatively large timebase (200 ms/div) had to be used. Using eq. 4.5, it can be seen that this timebase corresponds to a sampling frequency of 256 Hz, when the minimum sampling rate according to the Nyquist criterion [Ope89] should have been 363.64 Hz. The majority of the interferograms were sampled at higher frequencies, without any aliasing.

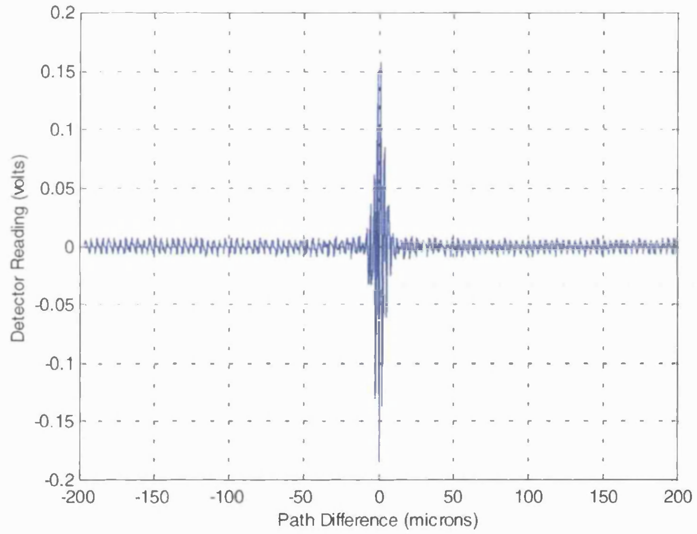


Figure 4.6 - Interferogram from a 20 W tungsten halogen lamp

The path difference scanning technique was similar to the one employed in FTS [Cha79], where recording a long interferogram can take considerable time, 4 seconds at the 50 μm/sec stage speed. Attempts to average many interferograms to increase the SNR were unsuccessful because the scan time was much longer than the reciprocal of the noise bandwidth, which made the signal appear at different instants in successive scans, adding incoherently and being averaged out. Another interesting feature of the interferogram signal displayed in fig. 4.6 is its rapid decay, as observed in fig. 4.2 a). The narrowness of such an interferogram is very useful in low coherence interferometry (reviewed in chapter two), as it yields a very high spatial resolution.

Fig. 4.7 shows an interferogram from the same filter used in fig. 4.3. As happened with the lamp, much more noise, seen as fluctuations in the peak amplitudes, is recorded with the fringe signal. This even created difficulties in finding the ZPD using the maximum amplitude (peak-to-peak). This problem is corrected in the coherence length calculation algorithm presented in the next section. Similar interferograms were measured from all seven interference filters available, to calculate their coherence lengths, and to observe the intensity of the interferogram sidelobe, essential to the IPSS technique. One of the filters was rejected for not having such a sidelobe, due to a smoother roll-off characteristic.

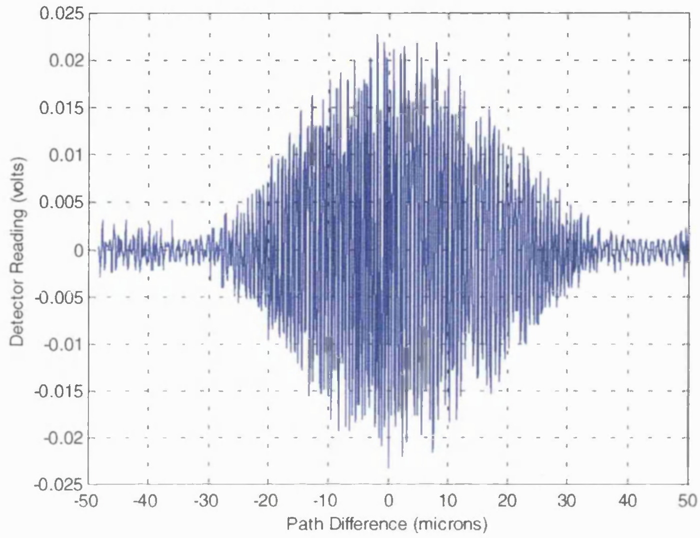


Figure 4.7 - Interferogram from an interference filter, mean wavelength 632.8nm, FWHM 11nm, measured with a translation stage

A variable coherence length source was implemented using the tungsten halogen lamp and a dispersive monochromator (Bentham M300), in the same manner implemented by Sutton [Sut82]. In a dispersive instrument, based on a single diffraction grating, the radiation collected at the exit has a colour (wavelength) that changes with direction, whose bandwidth is limited by a slit placed at the exit. By varying the width of this slit, it is possible to achieve different bandwidths, and hence, different coherence lengths. Provided that the light from the optical source fills the entrance slit, and that both monochromator slits have the same width, the bandwidth can be calculated to be [Ori00] the product of the grating dispersion, and the slit width. This source was used in studies of the effects of variable target-to-filter bandwidth ratios in IPSS, presented in chapter five. Figure 4.8 presents three interferograms from the Bentham monochromator, illuminated by a tungsten halogen light source with a blackbody temperature of 3220K. The slit sizes are 0.5 mm (a), 1.67 mm (b) and 5 mm (c). Considering the nominal dispersion of the monochromator grating of 1.8 nm/mm, the corresponding bandwidths from a) to c) are 0.9, 3 and 9 nm. The coherence lengths decrease, as expected, and are calculated in the next section. The a) plot has a poorer SNR, due to the lower light throughput of the instrument with a narrow slit setting.

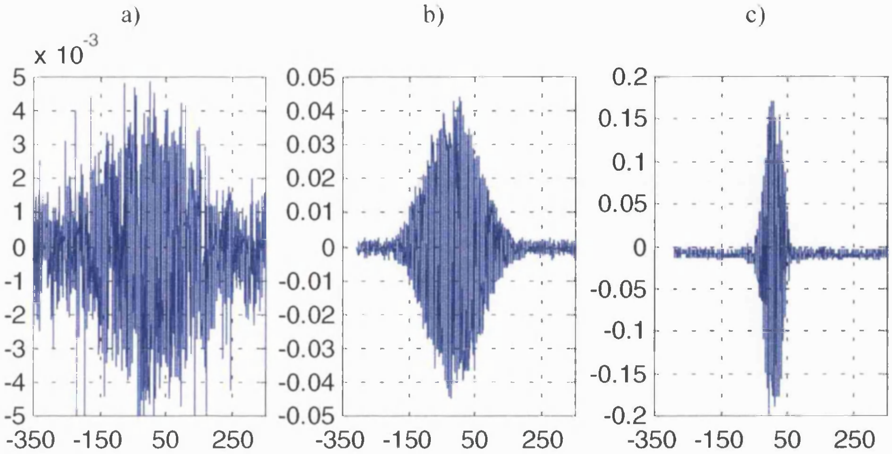


Figure 4.8 - Monochromator interferograms for different slit widths: a) 0.5 mm; b) 1.67 mm; c) 5 mm. The vertical axis is the detector voltage (volts); the horizontal axis is path difference (μm).

A fast piezoelectric transducer is used in chapter five to obtain interferograms with less noise, and with fringe carriers far from the noise spectrum. As happened with the first transducer, the path difference range is limited, allowing the measurement of full interferograms of white light only. One such measurement is shown in fig. 4.9, taken with a mirror scanning frequency of 300 Hz, mirror speed of 21.6 mm/s, and a fringe carrier frequency of approximately 78 kHz.

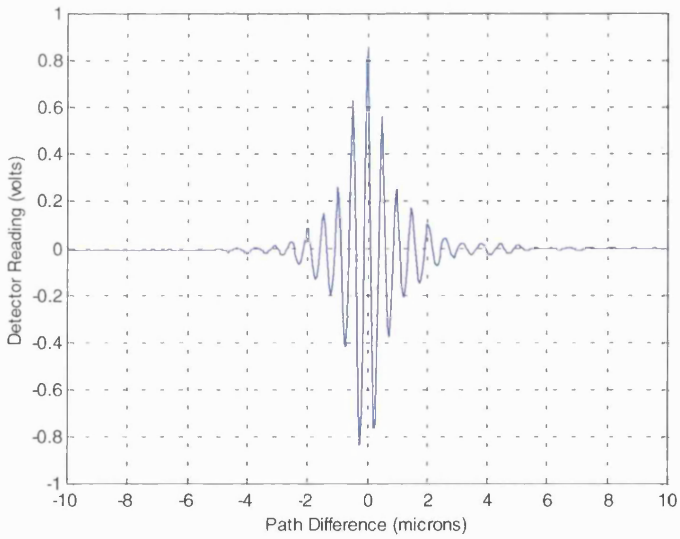


Figure 4.9 – Tungsten halogen bulb interferogram taken with a fast piezoelectric transducer

The plots included in this section are just examples of interferogram recordings performed in the course of the work. Many other interferograms were recorded using either the two piezoelectric transducers (one slow and one fast), the translation stage, or the imaging arrangement described in chapter six, such as from a laser diode, a He-Ne laser, or from resonant-cavity light emitting diodes (RCLEDs). The latter are presented in section 4.3.

4.2. Coherence length measurements

Measurements of coherence length find application in the characterisation of optical sources [Oul01] and in the assessment of the effects of light propagation in air or water. Swanson and others [Swa93] presented a brief review of coherence length measurement methods, and studied their suitability to the measurement of temporal coherence loss of laser light in water. The study showed that different coherence-loss mechanisms can be sensed by different methods, and that the direct (interferometric) method, besides being the simplest one, measures the effects of all mechanisms indistinguishably. This interferometric method consists in moving one of the interferometer's mirrors longitudinally, recording the fringe signal using a photodetector, digitising the resulting electrical signal, and processing the stored interferogram using a computer program. As presented in chapter 3, although the precise definition of coherence length is not a consensus among authors in optics, this concept is associated with the width of the self-coherence function, or with how far from the ZPD interference can be observed. In this section, coherence lengths are measured using the same experimental arrangement of the previous section (direct, interferometric).

The computer program, implemented in Matlab and listed in Appendix I, calculates the coherence length using two different formulae (eqs. 3.6 and 3.7). After loading the interferogram, it removes any residual DC level, takes the absolute values and allows the selection of a section of the interferogram, through the application of a boxcar window. Low-pass filtering of the interferogram signal is required in order to remove the fringe carrier, an additional modulation due to the translation stage (described later) and to obtain the amplitude of the envelope, which is proportional to the light source self-coherence function. The ZPD is assumed to be the path difference corresponding to the largest fringe amplitude. The width of the filtered waveforms is then assessed using the two formulae already mentioned, converted from sample numbers to μm using eqs. 4.1 and 4.5. A set of light sources, most of them used in the IPSS experiments, was chosen for coherence length calculation, comprising the tungsten halogen bulb, the ultrabright LED, the interference filters and the monochromator from the previous section, plus a laser diode from a laser pointer.

The most critical step of the measurement process is low-pass filtering. The electrical spectrum of the studied light sources was calculated using an FFT routine, included in Appendix I, to provide an indication of the cut-off frequency to be used in the software filter. This frequency was adjusted in each case to remove the fringe carrier and high-frequency noise, while maintaining the low-frequency envelope. An example of such adjustment is given in fig. 4.10, which displays an interferogram from an interference filter, and its corresponding self-coherence function, or interferogram envelope. In this case, the cut-off frequency was 20 Hz, for a sampling frequency of 512 Hz.

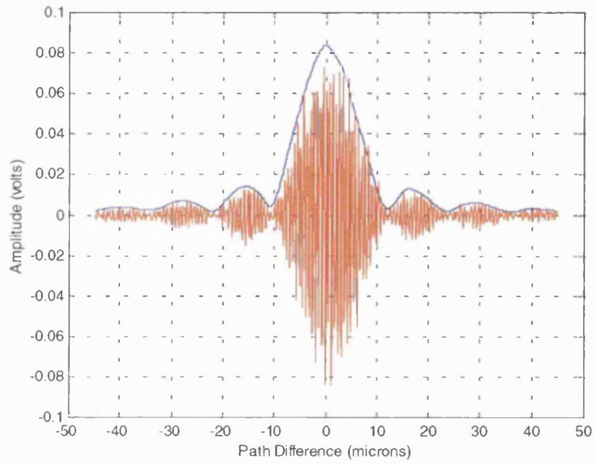
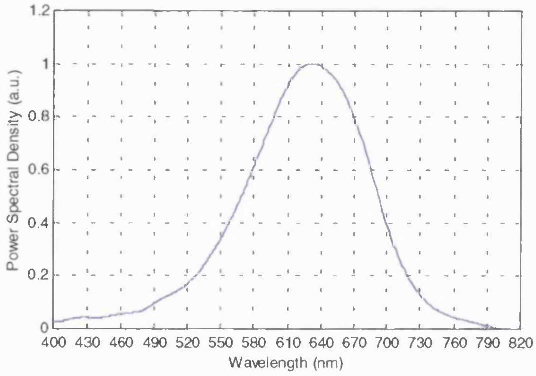


Figure 4.10 - Interferogram from an interference filter, mean wavelength 651.7 nm, FWHM 36.2 nm, and its filtered envelope

As secondary products of the spectral estimation process performed to design the lowpass filters, optical spectra of the same light sources were obtained by re-mapping of the frequency axis of the spectra to a wavelength axis. Fig. 4.11 presents two representative spectra, before filtering. The left one, from the tungsten halogen bulb, provides information about the optical bandwidth of the combined set of elements consisting of light bulb, interferometer and photodetector. The source spectrum does not vary considerably within the 400-800 nm range (6% in the Planck blackbody equation for a temperature of 3220K), thus the spectrum shown roughly indicates the response of the detection system composed of interferometer and detector. The variation of this response is negligible within the optical bandwidths of the interference filters employed in IPSS. The spectrum on the right plot of fig. 4.11 shows a main peak close to the nominal mean wavelength of the filter, 674.8 nm, and has approximately the same FWHM (18nm). The sidelobes on both sides are due to electrical noise, as observed in the corresponding values on the frequency axis.



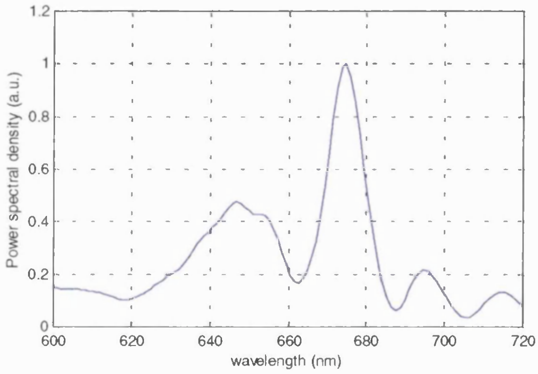


Figure 4.11 - Spectra from the tungsten halogen bulb (left), and from an interference filter with a mean wavelength of 674.8 nm, and FWHM of 18 nm (right).

The results of the coherence length measurements are presented in table 4.1. In all cases, the lowpass signal processing filter of choice was the Butterworth, 5th order, due to its maximally flat amplitude response [Ope89], and relatively simple implementation.

Light source	Sampling freq. (Hz)	Cut-off freq. (Hz)	Fringe freq. (Hz)	CL (μm) (Wolf) (eq. 3.6)	CL (μm) (Mandel) (eq. 3.7)	CL (μm) (eq. 3.3)	Fig.
Tungsten halogen bulb	1024	50	318	**1.16**	2.09	0.75	4.6
Tungsten halogen bulb with 651.7 nm interference filter, FWHM 36.2 nm	512	20	154	**5.66**	9.72	11.7	4.10
InGaAlP 644 nm LED, FWHM 18nm	512	20	147	**5.44**	9.32	23.04	4.2 b)
Tungsten halogen bulb with 632.8 nm interference filter, FWHM 11 nm	512	20	158	**12.59**	27.96	36.4	4.3, 4.7
Monochromator, wavelength 657 nm, dispersion 1.8 nm/mm, FWHM = 9 nm	256	5	150	**22.71**	45.28	47.96	4.8 c)
Monochromator, wavelength 657 nm, dispersion 1.8 nm/mm, FWHM = 3 nm	256	5	150	**57.60**	144.18	143.88	4.8 b)
Diode laser (from a laser pointer), wavelength 657 nm, FWHM 1.8 nm	256	2	150	**167.14**	209.28	239.81	Not incl.
Monochromator, wavelength 657 nm, dispersion 1.8 nm/mm, FWHM = 0.9 nm	256	5	150	**111.76**	209.27	479.60	4.8 a)

Table 4.1 - Coherence lengths of various sources, ordered by coherence length given by eq. 3.3

The columns of the table present: the sampling frequency used by the digital oscilloscope; the cut-off frequency selected in the low-pass Butterworth filter, a measure of the amount of noise included; the fringe carrier frequency, calculated using eq. 4.4; the coherence lengths calculated using eqs. 3.6, 3.7 and 3.3, and the figure in the previous section that presents the interferogram of the studied light source. As the monochromator and diode laser interferograms were very long, and the oscilloscope memory limited, and in order to cover the longer acquisition time required in these cases, the sampling frequency of the oscilloscope had to be reduced to 256 Hz. As mentioned in the previous section, this caused spectral aliasing in the acquired signals, which had little influence in the coherence length calculation, but impaired the spectral analysis. The coherence measurement is not affected since the algorithm uses the low frequency content of the signal, and the aliasing affects the signal frequencies above 128 Hz, half the sampling frequency. The interferogram carrier at 150 Hz was aliased down to circa 106 Hz, being filtered out anyway. Due to the problem with the spectral analysis, the fringe frequencies displayed are calculated, not measured.

It was observed that the cut-off frequency of the software filter had a strong influence on the coherence lengths estimated using Mandel's equation (3.7), and almost no influence when using Wolf's equation (3.6). Eq. 3.6 calculates the coherence length to be the RMS width of the self-coherence function, which does not vary with the presence or absence of the fringe carrier in the filtered signal. On the other hand, eq. 3.6 estimates the coherence length through the amount of integrated energy in the self coherence function, which increases as the dips in this function are filled by filtering out the fringe carrier. This behaviour led to the conclusion that Wolf's formula, although giving shorter coherence lengths than Mandel's, represents a better estimate of the coherence length from measurements in the presence of electronic noise. Hence, Wolf's formula was chosen to assess the coherence lengths in this study, as well as in subsequent studies by other workers [Gra01] [Oul01], being shown in bold in table 4.1. When compared to spectral measurements, the measurements of coherence length present the advantage of providing a good indication of the degree of spectral purity, independent of the mean wavelength of the studied radiation, thus allowing the comparison of this purity between radiations at different wavelengths.

The results in table 4.1 show that the measured coherence lengths have an inverse relationship to the optical bandwidths, as expected and mentioned in chapter three. There is no reasonable agreement between the measured quantities and the coherence lengths estimated using eq. 3.3, and this is due to the limitations of this estimate, discussed in chapter three, which considers a factor of unity in the reciprocal relation between spectral width and coherence length. The validation of the coherence length measurement method comes from comparison with indirect measurements done spectrally, shown in the next section. The source with the smallest coherence length is the light bulb, followed by the LED, interference filters, monochromator (with longer coherence lengths for narrower slits), and laser diode. In the case of the monochromator with a slit width of 0.5 mm and of the laser diode, the interferograms did not fit completely into the acquisition path difference range, rendering coherence estimates shorter than expected from the optical bandwidth. The coherence length of the monochromator with the narrowest slit should have been longer than the one from the laser diode, but using Wolf's equation (3.6) was not, because in the latter case a single-sided interferogram was acquired, flipped horizontally, spliced and put into the algorithm for calculation. Another interesting exception occurred in the case of the LED, which has displayed a coherence length much shorter than expected from its nominal FWHM. The device presents the same behaviour in the experiments in chapter five, when it is used as a simulated target in IPSS, and produces smaller shifts than expected. The reasons for this behaviour are not known.

4.3. A case study of resonant cavity light emitting diodes (RCLEDs)

Resonant Cavity Light Emitting Diodes (RCLEDs) are devices that employ technology from both semiconductor lasers (the resonant cavity) and conventional LEDs (the light-emitting material). One main difference is the active region, where spontaneous rather than stimulated emission occurs. The Experimental Solid State Group of the Centre for Electronic Materials and Devices, Imperial College (IC), has active research in this area, designing and manufacturing devices with diverse characteristics [Ico01]. In this section, work done in collaboration with that group is described, which comprised the investigation of the coherence length of RCLEDs, and its variation with the numerical aperture (NA) of the light collecting optics. The devices, fabricated at IC, are described elsewhere [Sta99], and their envisaged applications are low cost optical transmission systems based on plastic optical fibres [Gra01], and photodynamic therapy [She99]. A peculiar characteristic of these devices is the lack of cross-spectral purity, or the transverse variation of the spectral content of the light across the beam. This characteristic motivated the joint research described in this section and in published papers [Gra01] [Oul01] [Cou02], which comprised the assessment of the variation of the coherence length (and hence, the spectral linewidth) with the numerical aperture (NA) of the collection optics. This phenomenon was investigated using two different methods: UCL (the author) has measured coherence lengths, and IC has measured spectral bandwidths, and the variation of both with numerical aperture. Using the formulae in chapter three, the results were compared and found to be in good agreement.

The novel experimental arrangement used in this research consists of an adjustable numerical aperture coherence measurement system, presented in fig. 4.12, which has evolved from the arrangement used in the previous section. Coupling optics (which allow the input system NA to be controlled) followed by a scanning Michelson interferometer with an output silicon photodetector were used to record interferograms from 650 nm RCLEDs. One interferometer mirror was moved axially at 50 μm/sec by a motorised translation stage. The complete scan of 150 μm corresponds to a path difference range of 300 μm, large enough to measure the expected coherence lengths, below the 167 μm figure obtained for a semiconductor laser in the previous section. The electrical signal (interferogram) generated from the photodetector was amplified in an audio amplifier (not shown) and sent to a digital oscilloscope with a sampling rate of 256 Hz. AC coupling to the oscilloscope removed the DC background. The signal was transferred from thence, to a computer via a GPIB interface. The coupling optics from the RCLED to the interferometer consisted of a 45x, 0.65 nominal numerical aperture (NA) microscope objective, a 50 mm diameter, 62.9 mm focal distance plano-convex lens, and a 50 mm diameter iris. The radiation exiting the microscope objective had a 1.5 degree divergence due to the finite lateral extent of the source. A plano-convex lens was placed after the microscope objective to give a weakly convergent beam, and the separation of the microscope objective and the source was adjusted to give a spot size on the detector that matched its diameter of 10 mm. So, the RCLED aperture was imaged through the interferometer to the photodetector. This ensured that all of the radiation within the microscope objective's NA arrives in the detector. A variable NA is obtained by inserting a variable iris between the plano-convex lens and the interferometer. This allowed through only the central part with a diameter varying from 1.5 mm to 25 mm. The small source was positioned in the focal plane of the objective to give nearly collimated light out. The iris was widened to the edge of the beam and this was taken to correspond to the measured NA of the lens, 0.526. The NA measurement procedure is described later. Assuming that changes in the exit pupil are proportional to changes in the entrance pupil of the objective, then the NA can be calculated, and this has the same effect of placing a virtual iris in between the RCLED and the microscope objective (fig. 4.12). Fig. 4.13 depicts the spatial distribution of the light emission from the device. It was acquired with a CCD camera replacing the detector in the arrangement shown in fig. 4.12.

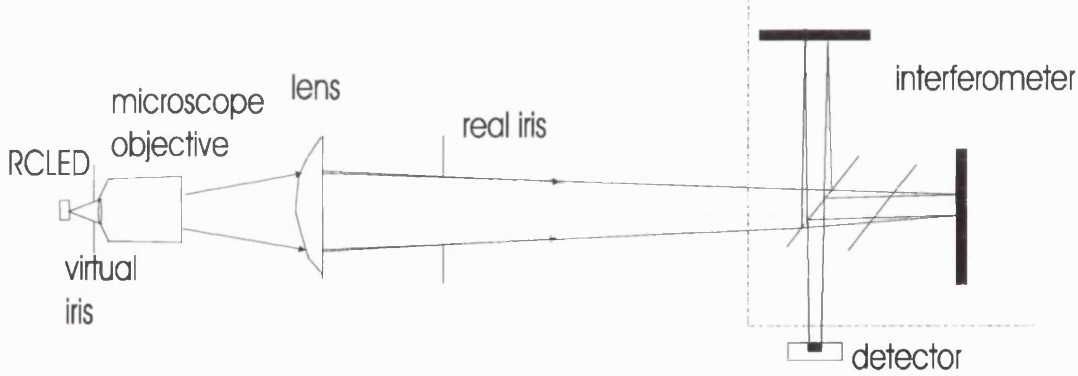


Figure 4.12 – Variable numerical aperture coherence length measurement system

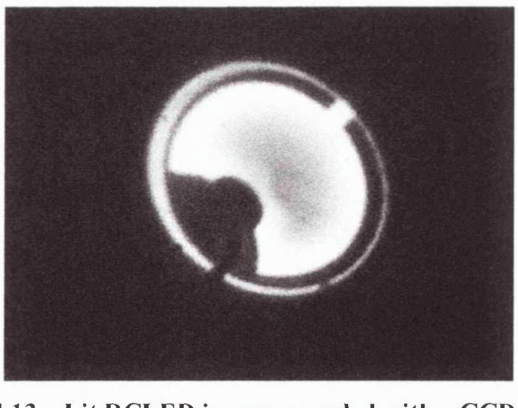


Figure 4.13 – Lit RCLED image recorded with a CCD camera

A total of 175 interferograms with 1024 samples each were acquired and processed, and a sample is presented in fig. 4.14. The signal processing was implemented by the program described previously. The software filtering removes both the fringe carrier and a modulation artefact found to be due to the stepper motor driving the motorised translation stage, slightly visible in fig. 4.14. This artefact is the consequence of variations in the mirror speed caused by the stepwise displacement of the translation stage.

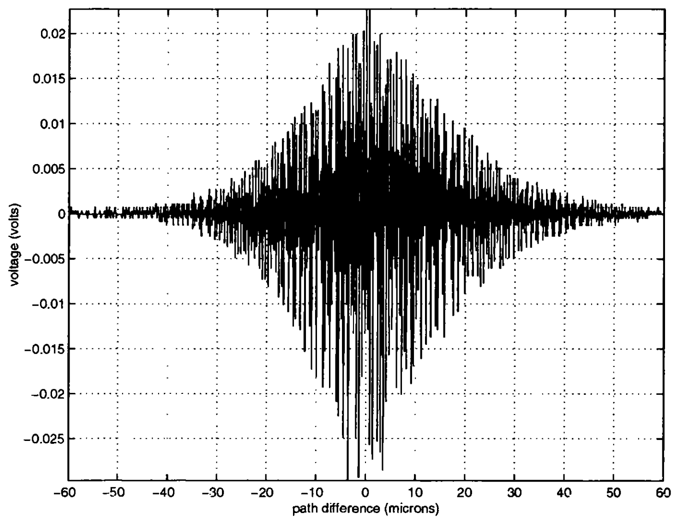


Figure 4.14 – Example of an interferogram acquired from an RCLED

Two of the four mesa diodes in the package were measured, one 400 μm and the other 200 μm wide. The 400 μm diode was measured with supply currents of 5, 20 and 40 mA, and the 200 μm diode with 5 and 20 mA. Seven different NA's were measured for each current in each diode, and for each NA five interferograms were recorded and processed by the computer program. The coherence lengths of each set of five interferograms were averaged and the standard deviation error bars plotted. The main effects of the imperfections in the experimental system components were the modulation effect coming from the mirror driving scheme, removed by software filtering, and NA uncertainty coming from the lack of knowledge of the real NA of the employed microscope objective, although other effects are discussed later in this section. To circumvent this last problem a simple NA measurement experiment was set-up, passing a He-Ne laser beam though the microscope objective, mounted on a translation-rotation stage. The objective was rotated and the size of the image on a screen was recorded. The image was not suddenly cut-off, but rather was gradually reduced in a peculiar roll-off from a maximum size to zero. As the iris diameter was visually adjusted so that the beam was tangent to its edges, the uncertainty in the process was characterised by taking the average and standard deviation of all observed image sizes, then interpolating this in the curve to obtain an average NA, the result being an effective NA of 0.5256, with a standard deviation of 0.0715. This was then taken to be the effective NA of the microscope objective. Further evidence of this roll-off effect is given by the fact that, within statistical error, no difference was found in coherence length between a point at maximum NA, and another point taken at 0.456. This means that from this NA upwards very little additional light is collected by the objective, thus not changing the coherence length.

As a part of the research collaboration scheme, the coherence length was also measured at Imperial College indirectly, using standard spectral techniques based on a monochromator with a spectral resolution of 0.1 nm [Gra01] [Oul01]. The device was mounted on a rotation stage, so that its spectrally resolved emission profile was measured for a range of angles and wavelengths, at a very small numerical aperture, determined by the size of a pinhole placed in front of the monochromator. This approach sampled the RCLED radiation in a small circular cell, off the axis normal to the surface of the device. Integration of the spectrally resolved radiation collected by this cell in a circle around the mentioned axis gives the spectrum in an annulus corresponding to an element of numerical aperture. The angular profiles measured at discrete angles are then interpolated using a Lorentzian curve, and integrated from zero to a particular angle, to give the spectral shape for that NA. The calculated spectra are then inverse Fourier-transformed to produce a self-coherence function, from which the coherence length is calculated using eq. 3.6. It was found out [Oul01] that the Lorentzian is the curve that best fits the spectral data, which is reasonable for a cavity-based device with a limited number of transverse modes [Sta99].

Fig. 4.15 presents two sample filtered interferograms for the minimum and maximum recorded NAs, clearly showing the curve broadening effect coming from the increase in coherence length as the NA decreases. The results of variation of coherence length with NA for the 400 μm diode, at 20mA, are shown in fig. 4.16. The two indirect curves are obtained under the assumptions that the lineshape is Lorentzian, or that it is Gaussian. The point obtained with an iris diameters of 25 mm (corresponding to the effective NA of 0.5256) is not shown, because the result is an artefact due to a limitation of the microscope objective, as previously explained. In the remaining six points, the two curves are in very good agreement, all points using the direct measurement lying in between the two indirect curves. In all curves, the coherence length decreases with increased collection angle or NA, as the spectral linewidth broadens [Gra01] [Oul01]. The direct measurement curve has stronger correlation with the Lorentzian curve for low NAs, and with the Gaussian curve for large NAs. It is believed that this change is due to the fact that, for small NAs, the device has a Lorentzian lineshape, as very few cavity modes are being measured, and, as previously mentioned, a Lorentzian lineshape would be expected from a cavity-based device. As the NA broadens, more transverse modes pass through the iris, and, according to the Central Limit Theorem, they combine into a Gaussian lineshape. Although Oulton and others [Oul01] have indicated that the best interpolation fit occurs for a Lorentzian, more accurately the spectral lineshape is a Voigt profile [Goo85], which is the convolution of the Gaussian and Lorentzian profiles, and their relative weights gradually change with NA. The directly measured point with the smallest NA has large error bars, which are due to the poorer signal-to-noise ratio arising from the small iris diameter (1.5 mm).

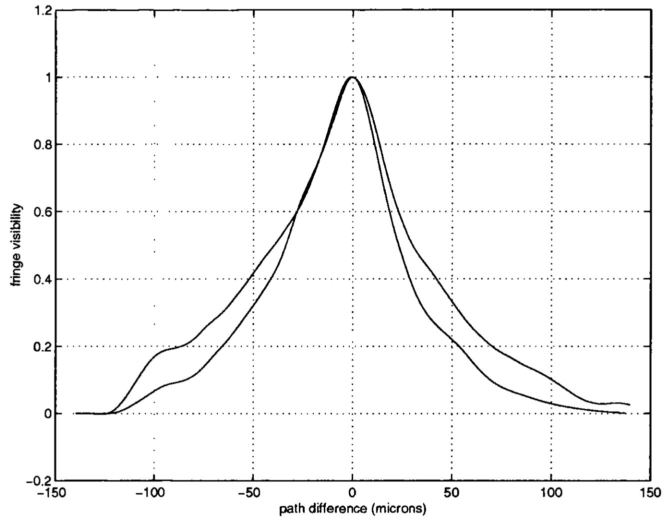


Figure 4.15 – Comparison of filtered interferograms acquired with the largest (inner curve) and narrowest (outer curve) NAs.

The coherence length decreases as the current increases from 5 mA to 40 mA, by 3% for low NAs and by 6% for high NAs. This corresponds to an increasing linewidth with current, which is reasonable as the increased number of carriers in the device's active region would increase the variance of the photon generation process. Since each device aperture was imaged to the full area of the photodetector, no difference was found between the 400 μm and the 200 μm diodes for the same NAs and currents to within experimental error. From this finding it can also be said that the non-collimation of the light from the source (1.5°), coming from its finite spatial extent, did not affect the measured coherence lengths, as the effects of extra path differences created by poor collimation, described in [Cha79] and in chapter three, were negligible. The measurements using different currents and device were not compared with spectral measurements, as these were not available. It is also important to point out that eq. 3.6 was employed under the assumption that the light field from the RCLED presents cross-spectral purity, which is not the case when the spectral contents of the light varies transversally across the beam. Although, due to the good agreement between the results using two different methods, this assumption did not cause large errors, further investigations of its effect on the calculated results need to be carried out.

There are other effects that usually impair the recording of long interferograms, which are mentioned in FTS literature, e.g. [Cha79]. Any eventual misalignment between the two interferometer mirrors, progressive with path difference, tends to reduce fringe visibility towards the edges of the interferogram. This effect was found to be negligible in the experiments in this chapter by measuring the visibility loss of fringes from a He-Ne laser, and deriving a correction factor that had little influence on the results. Another effect is the phase uncertainty caused by the lack of knowledge of the exact ZPD position. This usually happens when using one-sided interferograms, and requires the use of phase correction techniques [Cha79]. As all the interferograms acquired in this study were double-sided, and as the algorithm measures the full, not the half width of the interferogram, this effect was considered to be negligible.

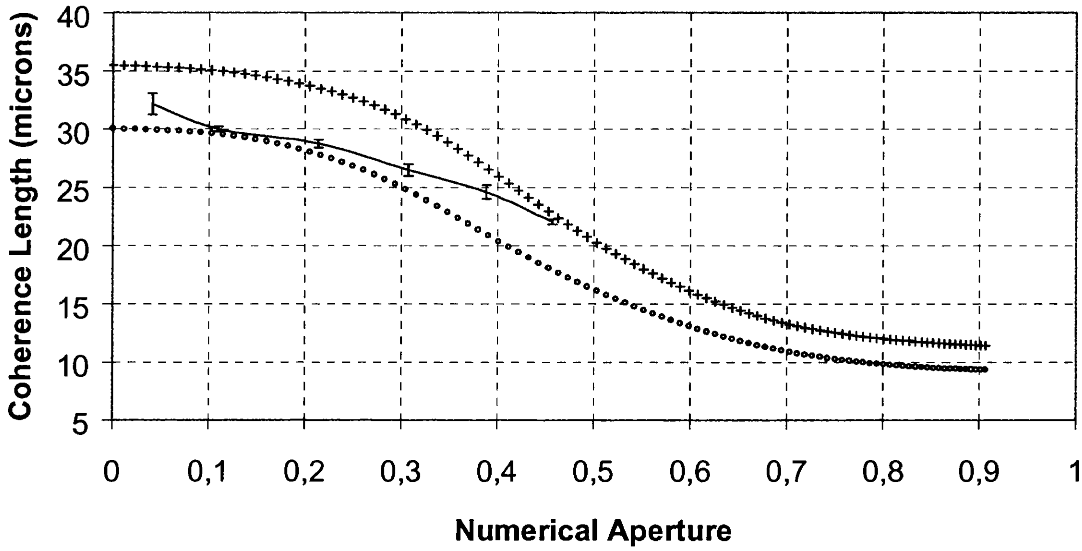


Figure 4.16 – RCLED coherence length measurement results, using Wolf's equation (3.6): upper dotted line – indirect measurement, Gaussian profile; lower dotted line – indirect, Lorentzian profile; solid line with error bars – direct measurement.

4.4. Conclusion

This chapter has presented the measurement of coherence lengths using a direct, interferometric method, which has the advantages of simplicity and of assessing all mechanisms that cause spectral spreading at once [Swa93]. The measurement process encompasses the recording of full interferograms and the processing of these interferograms to assess their widths. The former task gave a good insight of the problems encountered when performing real interferometric measurements, which allowed a better assessment of the relevance of the same kind of effects in the recording of short interferograms, as happens in IPSS. Another conclusion of the study is that Wolf's formula for coherence length (eq. 3.6) presents greater immunity to electronic noise, at least with the signal processing employed in this case. The coherence length measurement method was validated through comparison with independent spectral measurements, and this validation produced detailed information about the coherence properties of the light sources that are used in the main experiments of the thesis, through a metric of spectral purity that does not depend on wavelength (the coherence length).

The variable numerical aperture measurements of the coherence length of RCLEDs presented in this chapter allowed the characterisation of these devices with large collection angles, demonstrating the considerable variation of their coherence length (and linewidth) with numerical aperture, a concept rarely seen in the literature [Gra01]. The results produced are useful to the optical design of RCLED-fibre coupling, which may find application in cheap, short to medium haul optical links [Oul01]. Fig. 4.17 depicts the coherence lengths of the light sources measured in this chapter. RCLED coherence lengths (10-40μm) lie between ultrabright InGaAlP LEDs (5μm), and diode lasers (167μm), filling an important gap in the coherence spectrum.

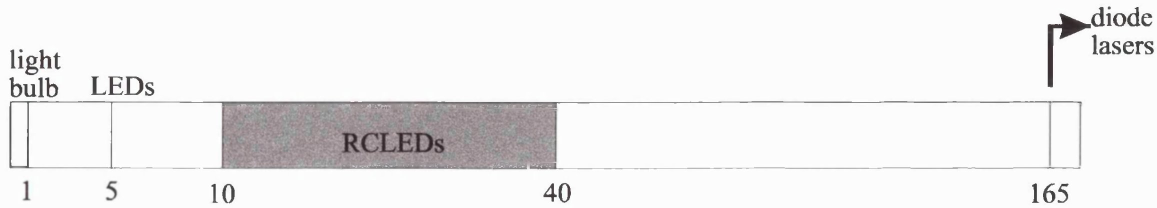


Figure 4.17 - Coherence lengths of artificial light sources

Chapter 5

Non-Imaging Detection Experiments

This chapter presents the most important experiments in this thesis. These were performed by changing the path difference in an interferometer through longitudinal displacement of one of its mirrors, and using photodetectors to convert an oscillating light intensity into an electrical signal in the delay-time domain (interferogram). Three Interferogram Phase Step Shift (IPSS) arrangements were employed, one with a translation stage, and two with piezoelectric transducers. The experiments with the translation stage and slow piezoelectric transducer had similar mirror scanning frequencies, around 2 Hz, allowing the comparison of their results to assess the behaviour of the mirror-driving scheme. The fast transducer experiment was performed to investigate whether the reduction in noise caused by higher scanning speeds could yield higher detection sensitivity. In the experiments described in the first section, the mirror is moved by a translation stage, controlled by a micropositioner, itself programmed by a computer. Besides some calibration runs, the section displays detection experiments of light sources studied in chapter four, investigations of the optimal target-to-filter bandwidth ratio, and of the effects of target-to-filter central wavelength offset. It is shown that very high sensitivities can be achieved when detecting a laser source, and high sensitivities to partially coherent sources, when compared to conventional approaches. These results can be further refined by choosing an appropriate target-to-filter bandwidth ratio. In another section, the two sets of experiments performed with a piezoelectric transducer changing the path difference are described, one with a relatively slow device, another with a very fast transducer, recording fringes at high frequencies, for interferometric applications. It is shown that the translation stage scheme, although suffering from higher noise, produces more accurate results, being more suitable for theoretical studies, while the fast piezoelectric transducer arrangement is more adequate when one needs higher target update rates. The results displayed throughout the chapter were obtained with an algorithm whose selection and detailed description are given in chapter seven. A comparison of the main experimental results with the theory is included in chapter eight.

5.1 Experiments using a translation stage

5.1.1. Description, alignment and adjustments of the experimental system

The basic experimental arrangement employed throughout this chapter is presented in fig. 5.1. Further versions incorporating minor modifications are introduced in the next sub-sections. The detection system, shown within the dotted box in fig. 5.1, is an implementation, in the visible band, of the IPSS theoretical system described in chapter three. Although, as mentioned in chapter three, the system can be implemented in any band where components are available, this band was selected due to its ease of alignment. As happened in chapter four, the experiments were assembled on an optical table floated with nitrogen to reduce the effects of the environment. The elements outside the box simulate the radiation that reaches the detection system. In this arrangement, one light source represents a target, and can be either a He-Ne laser or one of the partially coherent sources studied in chapter four. Another source simulates the incoherent background, comprising a tungsten halogen lamp with a fixed blackbody temperature of 3220 K. Neutral density filters were inserted in the path of the radiation from the target, allowing a variable target-to-background power ratio to be achieved, as done by Sutton [Sut84] and Duffy [Duf89]. This ratio is from now on termed signal-to-clutter ratio (SCR). The light from the two sources combine in a non-polarising cube beamsplitter to enter the detection system, where a reference plane is defined in the dotted line above the interference filter, where power measurements are taken to establish the experimental SCR.

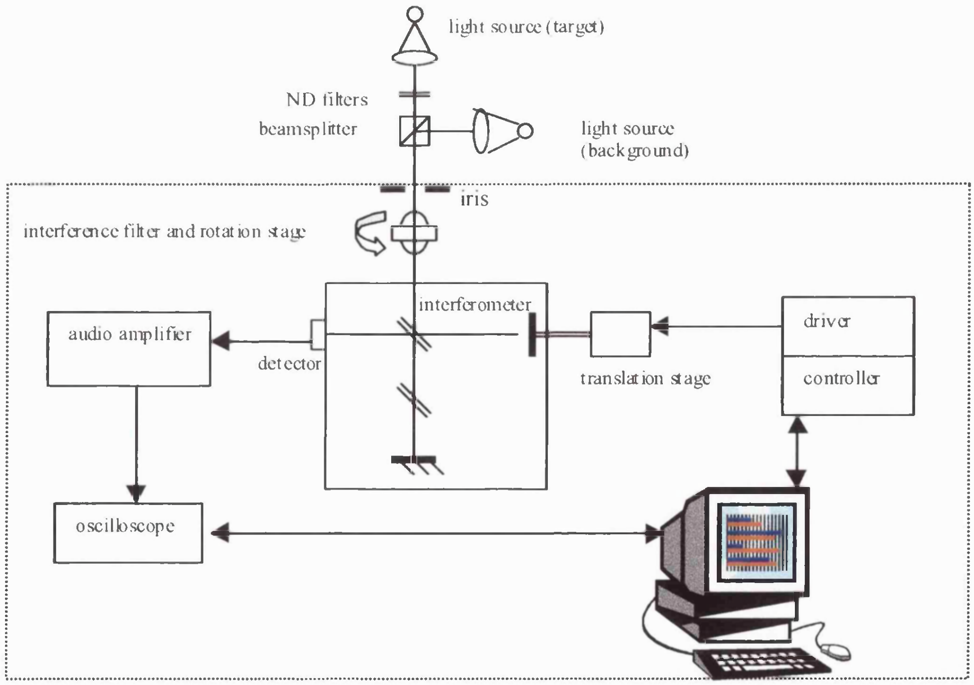


Figure 5.1 - Basic experimental arrangement employed in this chapter

The light exiting the beamsplitter is then spatially limited by an iris, whose purpose is to allow better control of the amount of target and background power that reaches the system. In each experiment, the iris diameter was set in an attempt to limit the beam diameter to that of the components and apertures of the system, so that all light measured at the system's entrance actually interfered and reached the detector. The critical points regarding this requirement were the 28 mm interferometer mirrors, which had the smallest diameter among the optical components in the beam path. The light that enters the detection system is filtered by an interference filter, which is selected in accordance with the central wavelength of the target to be detected. The filters employed are multi-layer devices that exploit the Fabry-Perot cavity effect [Hec98], and, because of the directional characteristic of these devices, were mounted on a rotation stage so that maximum transmission of the energy from the target could be achieved. This alignment is further discussed when the effects of the offset between target and filter central wavelengths are studied. The filtered beam is sent to a Michelson interferometer, the same used in the experiments in chapter four, which measures the self-coherence function of the incident light, for a range of path differences determined by the travel of one of its mirrors, moved by a translation stage. The set translation stage - motor controller - computer is the same used in chapter four, the difference being the short path difference range employed. As pointed out in chapter four, only a small region of the interferogram, close to its first minimum, needs to be scanned in IPSS. The interferogram section exiting the interferometer was collected by a photodetector that, in this section, is the same used in chapter four. Focusing optics were employed to collect the light from the exit aperture of the interferometer, and focus it onto the detector surface. The detector, operating in the photovoltaic mode, was connected to a digital oscilloscope (Tektronix 2430A), which removed the DC level from the converted signal and stored the AC-coupled interferogram sections as ASCII data files with 1024 samples each.

The region of the interferogram where the IPSS technique operates is also the one with the poorest signal-to-noise ratio (SNR), due to its lower fringe visibility. In order to obtain enough SNR in the recorded interferograms for off-line processing, an audio amplifier (EG&G Brookdeal 9452) was employed to amplify the fringe signal from the detector. The amplifier also had a built-in bandpass filter, whose low and high cut-off frequencies could be set on its front panel. Another useful tool to increase the SNR was signal integration, by averaging interferograms acquired in successive scans. In order to guarantee that integration was done coherently, and that an adequate no-target reference was obtained, it was essential that the mirror scan and the oscilloscope acquisition were synchronised. This was done by triggering the oscilloscope with a signal generated by the motor controller at the start of each scan. Because the detection system operates in a repetitive fashion, scanning a limited region of the interferogram, the use of a reference laser for path difference calibration, as happens in FTS, is not required. This is an advantage of IPSS, which simplifies the hardware and reduces the implementation costs. The motor controller that moved the translation stage was itself controlled by a computer program in Labview language. In this simple program, it was possible to control the speed of the stage, the length of the scan, and the number of scans to be averaged.

This number was selected simply by observation of the scan screen, i.e., the lower the original SNR, the higher the number of scans averaged. A caveat about this procedure is the fact that exaggerated averaging removes the sought-after interferogram phase step, as it corresponds to a lowpass filtering operation, removing the high-frequency phase transition. This effect is further discussed in chapter seven.

A picture of a typical implementation of the system is shown in fig. 5.2. The red lines show the main beam propagation paths. Appendix II lists the interferograms acquired for IPSS experiments, presenting the values of the different parameters varied throughout the experiments. The table also shows experiments done with piezoelectric transducers (section 5.2), and imaging experiments (chapter 6). Some modifications were introduced during the course of the work, such as the iris at the entrance to give better control of power measurements, and a focusing lens at the interferometer exit, to increase the SNR.

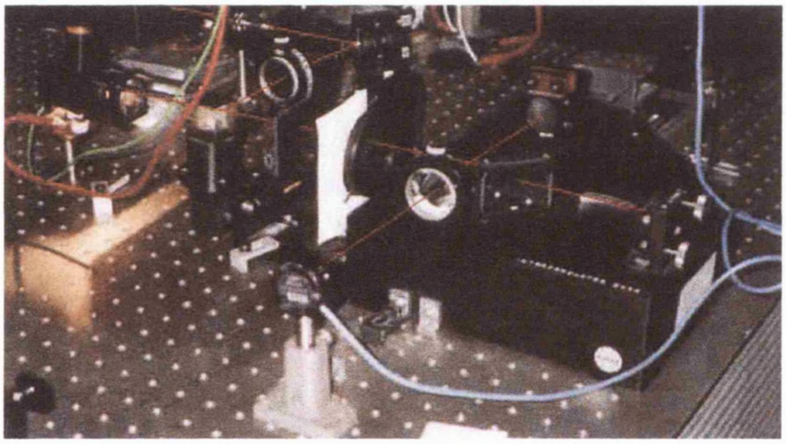


Figure 5.2 - Experimental arrangement

A typical interferogram section, displayed with its corresponding trigger signal, appears in fig. 5.3. The translation stage was moved at 1500 steps/sec, and the oscilloscope timebase was 20 ms/div. The falling edge of the trigger (a) is synchronised with the start of the scan; after a delay of approximately 35 ms, the stage starts to move and to produce fringes. This delay is due to the limited frequency response of the mechanical set comprised by the mirror, its supporting frame, adapter and translation stage; the noisy regions observed just after the edges of the trigger signal are caused by the random oscillations of the mirror when the stage is reversing; hysteresis can also be seen in the plot, as the two interferograms are not absolutely symmetric. This characteristic prevented the integration of both scan directions; only one had to be chosen. Another feature necessary in the detection algorithm is means of chopping off the edges of the interferogram, as they did not represent fringes, but rather noise from the stage reversal. If the algorithm cuts off all interferograms in a set at the same points, this does not affect the displacement measurement, which constitutes the basis of the technique.

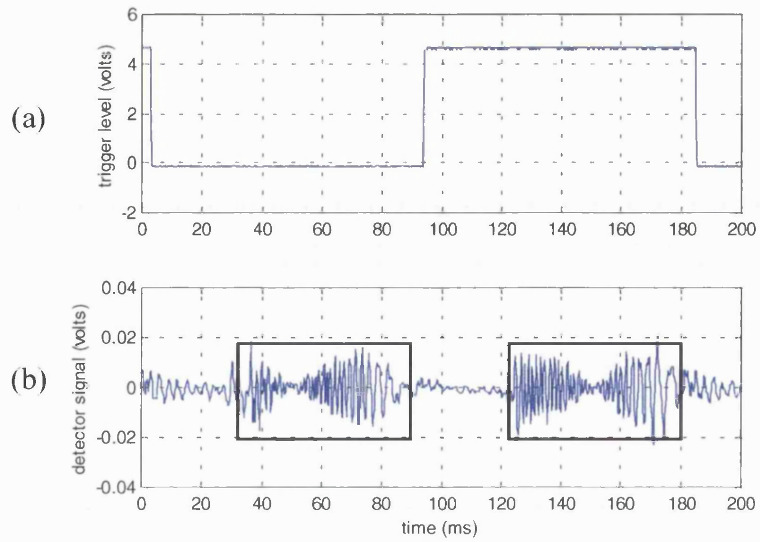


Figure 5.3 - Typical interferogram (b) and its corresponding trigger signal (a). The regions of interest are shown within the rectangles.

As described in chapter three, the sensitivity of the IPSS approach is governed by the smallest detectable shift of the interferogram phase step. This shift is the signal to be detected, and competes, as in any other detection system, with noise. In the IPSS case, anything that produces a displacement in the position of the phase step in the path difference axis, and is not due to the presence of a coherent target in the field of view of the system, can be considered noise. Consider the recording of an interferogram with the tungsten halogen bulb only. If the position of the amplitude minimum moves from one scan to the next, either the measured interferogram or the initial mirror position has changed. In the first case, either the tungsten bulb radiation changed, or the interference filter did. In the second case, the mirror driving mechanism may have moved from one scan to the next. Thus, the main factors causing such displacement are background spectral fluctuations within the filter band, changes in the interference filter response with temperature, and lack of repeatability in the translation stage. As mentioned previously, the IPSS approach does not require a reference laser, but does require good repeatability in the start of the path difference scan. Although the three effects could not be assessed separately, it is believed that the dominant effects are those from the tungsten bulb and the translation stage. The filter response changes with temperature, but the duration of the experiments (a fraction of a second) is much less than the usual temperature change cycle, thus being negligible. The tungsten bulb radiation does fluctuate 100 times per second (twice the supply frequency), and this amplitude fluctuation combines with the target strength, producing a displacement in the phase feature even if the target power is constant or absent. When a partially coherent target is present, its amplitude fluctuations, expressed by its variance, add to the background noise. If the target is a laser, the dominant noise source is phase induced intensity noise [Ari91], which is the conversion of laser phase noise into amplitude noise in the interferometer.

The translation stage has a finite repeatability, not stated in the stage documentation, but smaller than 0.1 μm, which is the stage resolution. Fig. 5.4 depicts a simple experiment performed to measure the repeatability of the translation stage. An interferogram from a tungsten halogen bulb, with a sharp amplitude maximum, is used as a path difference marker. The upper plot shows the initial ZPD position, where the positive amplitude peak occurs; a command is given to the stage to move 50 steps in one direction, and a second interferogram is recorded (middle plot). A command of 50 steps in the opposite direction is then given, and the mirror was supposed to return to the starting position.

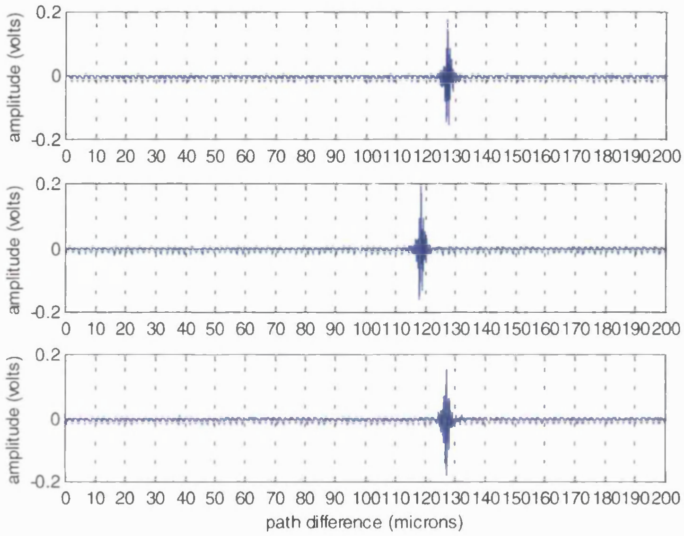


Figure 5.4 - Assessment of the repeatability of the translation stage

The position of the amplitude maximum in the lower plot is one sample less than in the upper plot. Using the conversion factor given in eq. 4.5, this corresponds to an error of 0.0977 μm, in reasonable agreement with the limit given by the stage resolution (0.1 μm in mirror position, or 0.2 μm in path difference). Another investigation was carried out with a different set of white light interferograms, to assess the accuracy of the conversion factor given by eq. 4.5. Four interferograms were recorded with commands of 100 steps in between them (corresponding to displacements of 10 μm each). The three measured displacements had an average of 9.80 μm, with a standard deviation of 0.64 μm, an error of 2%.

The high sensitivity of the IPSS approach comes from the use of coherence rather than intensity of the detected light, and from the use of phase, rather than the amplitude, of the coherence interferogram. The novel algorithm used to extract phase and frequency information from the interferogram is described in chapter seven. Basically, it applies a bandpass filter to the interferogram and calculates its instantaneous phase and frequency from the analytic signal [McD98]. In fig. 5.5 an example of the processing tasks of the algorithm is given. Two waveforms are seen in each plot: the upper with the target present; the lower without it. In plot

a) the signals are displayed as recorded; in b) they were filtered using a software filter; plot c) shows the unwrapped instantaneous phase, while plot d) shows the instantaneous frequency, which is the marker used to measure the displacement in the phase step.

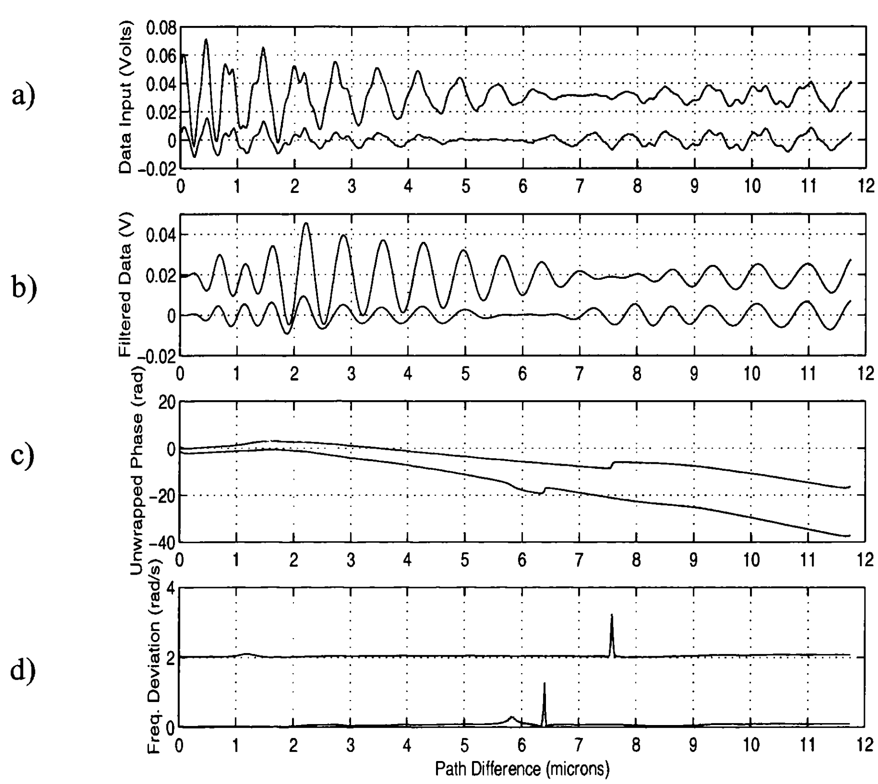


Figure 5.5 - Algorithm signal processing tasks. (a) - input signals, (b) - filtered signals, (c) – instantaneous phase and (d) – instantaneous frequency.

As previously mentioned, care must be taken when selecting the number of scans to be performed to record an adequate interferogram. This number influences the performance of the phase transition detection process. A low interferogram SNR, with too little averaging prevents the detection of the phase step, as it appears immersed in noise; too much averaging removes the high frequency content which contains the phase step, preventing the detection too.

Fig. 5.6 illustrates this behaviour. In the left plot, the red interferogram has not been averaged, and the blue interferogram was averaged three times. The phase transition and frequency spike are clearly discernible in the latter case. Averaging one time produced a very small spike. In the right plot, the red interferogram was averaged ten times, and the blue one thirty times. The amplitude of the frequency spike when averaging ten times is 1/20 of the corresponding amplitude when averaging 3 times; when averaging 30 times, the result is even worse. Thus, for the interferograms in the conditions seen above, averaging three times would give the best result (higher frequency spike amplitude), although this number does vary with the experimental conditions. Another noticeable effect is the phase distortion caused by excess averaging. The approximately linear, decaying unwrapped phase behaviour seen in the left plot of fig. 5.6 and in fig. 5.5 is no longer seen in the right plot. This linear phase is in agreement with the theoretical interferogram presented in chapter four (see fig. 4.5). Another resource to improve the interferogram signal-to-noise ratio is the bandpass filtering done by the computer algorithm, described in detail in chapter seven.

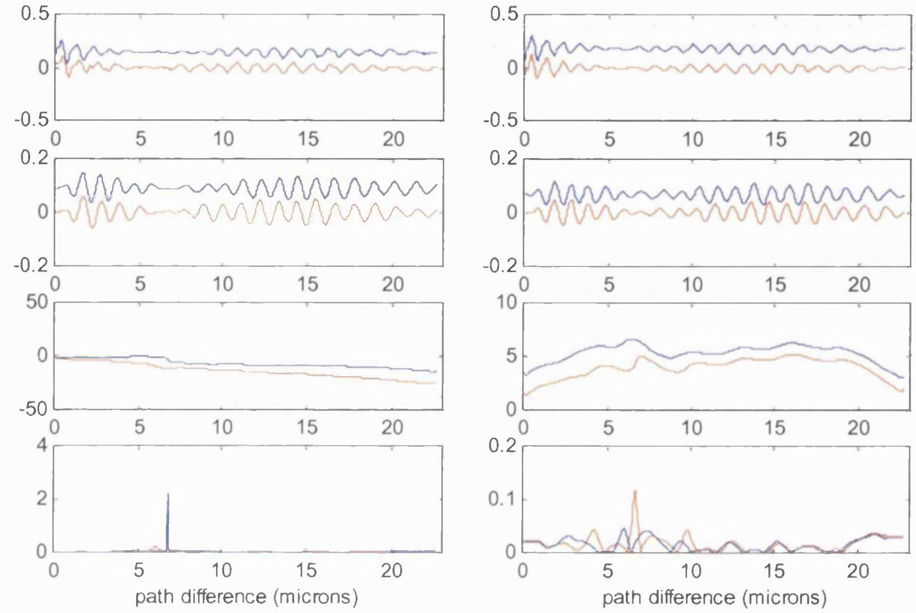


Figure 5.6 - Effect of averaging on the phase step. Left: no averaging (red), 3 times (blue). Right: 10 times (red) and 30 times(blue).

Due to the finite repeatability of the translation stage, and in order to minimise the effects of time drift, it was decided to collect the experimental data in pairs of interferograms, one with the target present, and the other without it. By using this procedure, the variation in conditions in between the acquisition of the target-present and target-absent interferograms is minimised, thus reducing the experimental error. Besides the acquisition in pairs, five points (interferogram pairs) per SCR were acquired and their average and standard deviation were calculated to characterise the dispersion of the measurements statistically. Although not practical in an application-oriented system, this approach was more adequate to our research purposes.

5.1.2. Measurements of phase step shift

After setting-up the experimental arrangement and characterising its repeatability, among other parameters, the system was employed to record the phenomenon under investigation. The first target to be measured was a 5 mW unpolarised He-Ne laser (Melles Griot 05 LLR 851). It was switched on and left working for one hour before any measurements were taken. The laser beam was expanded using a 10 times microscope objective, slightly tilted with relation to the laser optical aperture, so that the reflection of the laser beam did not return to this aperture. This was done to avoid increase in the laser Relative Intensity Noise (RIN) figure, caused by reflection from the backside of the objective. It was decided not to use spatial filters in the experiment, since the reduction in coherence length caused by laser transverse modes was not enough to alter the system's sensitivity to the laser, being such a filtering scheme of little value. Additionally, it was desirable to keep the experiment as simple as possible. According to the manufacturer [Mel99], the laser had a nominal linewidth of 0.002 nm, comprised of a set of four or five Lorentzian longitudinal modes propagating under a Gaussian doppler spreading curve, and its coherence length was approximately 20 cm. The expanded laser beam was passed through a double achromat lens to collimate it and combined in a non-polarising cube beamsplitter with light coming from a 20 W tungsten halogen bulb with a blackbody temperature of 3220 K. We have measured a beamsplitter splitting ratio of 45:55. The light bulb was placed in an enclosure with a concave rear mirror, a front aperture stop and a 30 mm f/2 plano-convex lens to collimate its beam. Neutral density filters were placed in the collimated laser path to produce twelve different SCRs. The light from target and background, combined in the beamsplitter, passed through a 632.8 nm centred, 11 nm FWHM interference filter and propagated into the interferometer to be focused into the detector surface (of 1 cm^2) by another plano-convex lens placed at the interferometer's exit aperture. The interferometer's moving mirror was scanned by the 0.1 μm step translation stage. Four scans were averaged to produce one signal. The computer program was set to scan the mirror position in a linear manner from one extreme to the other before it returns to its starting position and repeats the motion in a succession of ramps. The length of the scan, the scan speed and the number of scans are programmable. The measurements were taken using a 10 μm scan length, corresponding to a 20 μm path difference range, with a speed of 150 μm/sec in steps of 0.1 μm.

The electrical signal so generated was filtered and amplified by 30 dB in an audio amplifier. The speed of the scan allows the determination of the mean fringe frequency to be in the range 150 to 160 Hz for the wavelength range used in the experiments (630 to 680 nm), see eq. 4.4. The mirror scanned at around 2 Hz. Therefore, the analogue filter's 3 dB cut-off frequencies were set to 100 Hz and 10 kHz. A digital oscilloscope acquired this signal with AC coupling at a sampling frequency of 5.12 kHz, removing the DC level. The translation stage suffered from hysteresis so only one of the directions of the scan was used. Each 20 μm path difference range gave a data block of 1024 samples, which was then transferred to the computer using the IEEE-488 interface for off-line processing. An example of a data block is given in fig. 5.7.

The interferogram signals were recorded as previously described and processed by the computer algorithm described in detail in chapter seven. The filtered SCRs were measured by placing the same photodetector used to measure the fringes behind the iris shown in fig. 5.1, and recording the powers from target and background. The SCRs at the input of the optical system were measured at the same system input plane, by removing the interference filter. Target power ranged from 0.935 nW to 2.77 μW, while background power was in the range of 32.5 μW to 145.5 μW. Daylight power levels measured with the same detector in London, during summer, were approximately 3 mW at sunset with overcast sky, and 40 mW at noon, sunny weather. As the IPSS sensitivity is determined by the target to background power ratio, if the experiments demonstrate (as they actually did) that the approach works with background powers of a few μW, it would also work with mW power levels, and the interferograms would even have better SNRs, due to the higher background power.

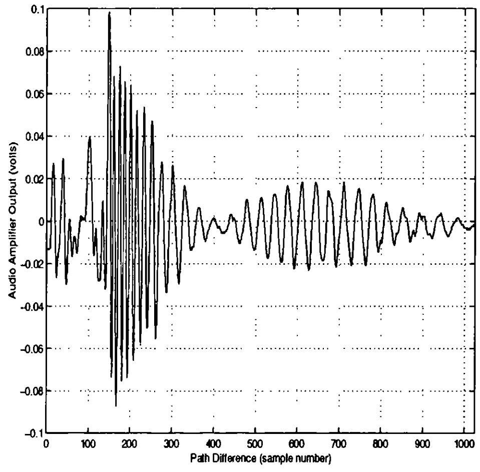


Figure 5.7 - An example of an interferogram

Fig. 5.8 presents the results of the phase step shift measurements. It can be seen that little or no displacement in the phase step happens until the filtered SCR reaches –30 dB, from where the curve clearly departs from zero and the laser is being detected. A simple definition of the system's sensitivity would be to measure the smallest SCR from which the error bars no longer touch zero [Cou99] [Cou00]. By using this criterion, the sensitivity of our detection system would be –27.74 dB after filtering, and –46.42 dB at the system's input, very high sensitivities when compared to most of the detection systems reviewed in chapter two, which is done in chapter seven. In chapter seven it is also shown that this criterion is actually of limited usefulness, as it does not encompasses the probabilities of detection and false alarm. In that chapter, the system's performance is characterised through the construction of its Receiver Operating Characteristic (ROC) curves. The average dispersion coefficient (standard deviation/average) in these measurements was 1.05, a figure used later for comparison. From the SCR of –8.5 dB an observability cut-off exists, as the target amplitude is so strong that the phase step no longer occurs. Chapter eight discusses this effect in the light of the advanced theory presented there, and the curve in fig. 5.8 is compared with the theoretical results.

Due to the finite size of the bulb's filament, the degrees of collimation of target and background were different. This had little impact on the results of the experiments, since the path difference scan was very short, and the influence of the extra path difference created by poor collimation (explained in chapter three) is proportional to the scan length [Cha79].

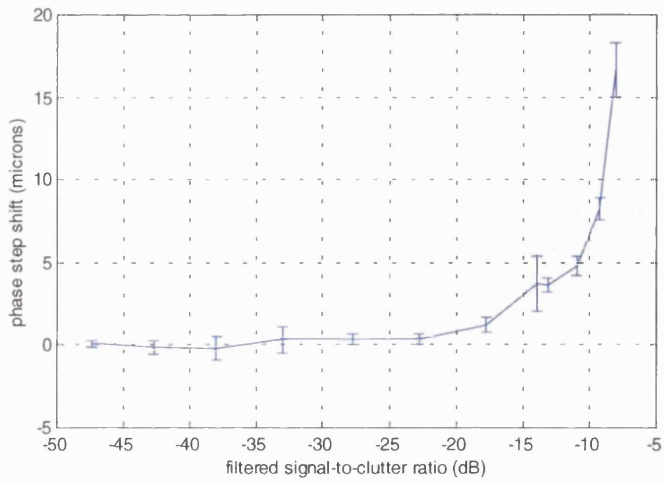


Figure 5.8 - Detection of a He-Ne laser

In order to assess the sensitivity of the detection system to partially coherent targets, another experiment was performed using two light sources: one to simulate the background, and the other to be filtered by an interference filter, simulating a target. A second 30 mm, f/2 plano convex lens was used to collimate the light from the light source. The target's filter had a central wavelength of 648.7 nm and a FWHM of 12.2 nm, while the detection system's interference filter had a central wavelength of 651.9 nm and a FWHM of 36.2 nm. The central wavelength offset was compensated by tilting the system's filter using the rotation stage. The target power was in the range 39.9 nW – 10.5 μW. Fig. 5.9 presents the phase step shift results obtained using this target. Much more noise than in the laser case is now present, which shows the considerable influence of light bulb intensity fluctuations with the 50 Hz supply, as, instead of one, two such light sources are employed. The increase in noise is reflected in the dispersion coefficient, in this case 1.7. Using the simple criterion previously mentioned, the sensitivity to this light source is –20.39 dB after filtering, or –31.96 dB at the detection system input. Even for partially coherent sources, the sensitivity of IPSS is still high, compared to other approaches, but it is lower than in the laser case, as expected from a coherence-based approach.

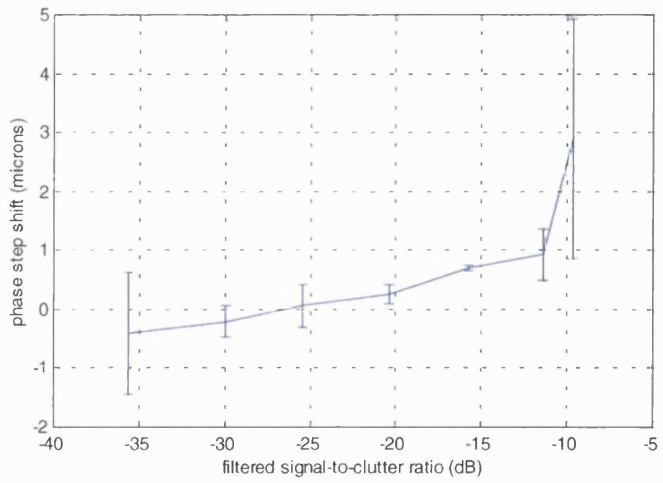


Figure 5.9 - Detection of a filtered white light target

In order to investigate the variation of sensitivity with target coherence, another partially coherent target was constructed using an ultrabright InGaAlP 15000 mcd LED, with a central wavelength of 644 nm and a FWHM of 18 nm (manufacturer data). An aperture stop and a 30 mm, f/2 plano convex lens were again used to obtain collimated light. The filter used in this case was the same as that used with the filtered white light target, and the rotation stage at the system entrance was used to maximise the LED power coupled through the interference filter. The target power was varied by changing the LED supply current, from 0 to 40 mA. Fig. 5.10 presents the measurements of the displacement in the phase step. The scales of the plot were changed to allow better visualisation. The average dispersion coefficient was 2.5, displaying the highest noise level of all measurements in this section. Using once more the simplified detection criterion, the LED was detected at a filtered SCR of –5.85 dB, corresponding to an SCR of –13.07 dB at the detection system input, the lowest of the three figures presented so far.

The resonant cavity light emitting diodes (RCLED) whose coherence lengths were measured in chapter four were also used in detection experiments, again for comparison purposes. Due to its peculiar cross-spectral purity characteristic, the RCLED mounting was implemented as done in the experiments in that chapter. The device was mounted on an optical rail containing a microscope objective with a nominal NA of 0.65, a plano-convex lens (30 mm, f/2) and an iris to control the NA, in this case adjusted to the maximum NA. As in the LED case, different SCRs were obtained by varying the device's supply current, from 0 to 40 mA. The microscope objective's working distance was adjusted to give an image of the RCLED at the detector plane, contained within the beam coming from the tungsten halogen bulb. Fig. 5.11 displays the results of the phase step displacement experiment.

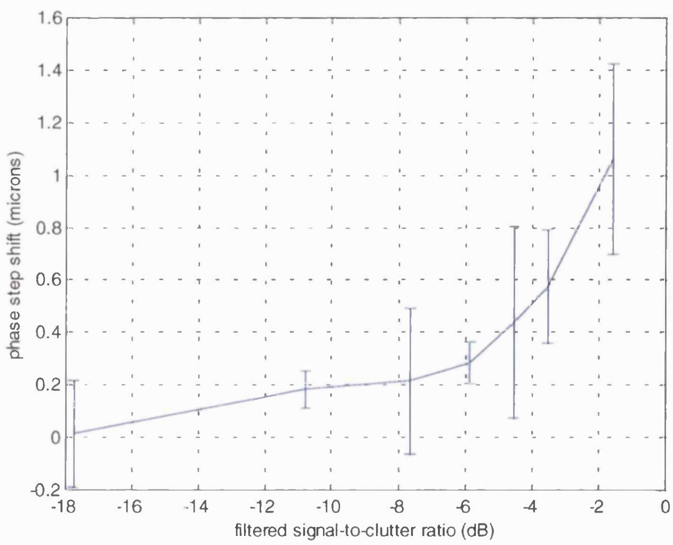


Figure 5.10 - Detection of a LED target

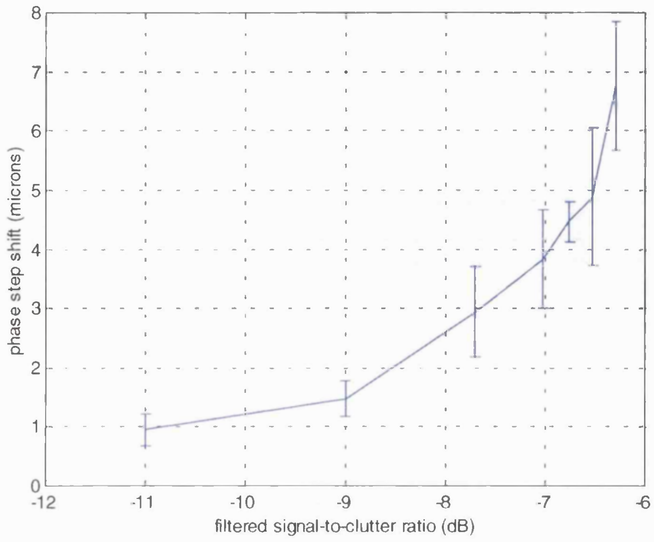


Figure 5.11 - Detection of an RCLED target

In this case exclusively, the SCR displayed on the abscissa of fig. 5.11 is the power ratio measured at the detector plane. This was done due to the lack of room to position the detector at the entrance of the interferometer, after the interference filter, to obtain a reliable power measurement, which would have guaranteed that all the power from both target and background would reach the detector. Target and background powers were then measured at the detector plane (interferometer exit), and the SCR at this point can be considered to be the same as at the interferometer entrance, if one assumes that optical transmission is the same for filtered target and background. This was measured and found to be true, due to the relatively narrow optical bandwidth of both light sources, after filtering by the interference filter, and to similar degrees of collimation. The measurement's average dispersion coefficient was 0.21, showing that the RCLED has a very stable output. In fig. 5.11 one can observe that, using the same simplified detection criterion as for the other targets, the RCLED was detected at the lowest of the SCRs measured, –11 dB after filtering, or –30 dB at the system's input. As lower SCRs were not used, the minimum detectable SCR may be even less than this figure. Target power levels were in between 20 and 67 μW.

The variable current used in the RCLED experiment had little influence on the result, as the coherence length (hence linewidth) of the radiation from the device varied only 3% with current, as shown in chapter four. In the case of the InGaAlP LED, the linewidth was measured by the manufacturer at a current of 20 mA, but its variation with current was not known and may be one reason for difference between experiments and theory, discussed in chapter eight.

Table 5.1 presents a summary of the phase step shift results obtained with the light sources investigated in this sub-section. Target central wavelength and FWHM data was reproduced from the manufacturers, apart from the RCLED, whose linewidth was taken from chapter four [Oul01]. The reported coherence lengths were measured using the method described in chapter four and calculated using eq. 3.6 , with the exception of the laser, given by the manufacturer.

Target	Target central wavelength and FWHM (nm)	Target coherence length	Filter central wavelength and FWHM (nm)	Dispersion coefficient (standard deviation / average)	Minimum detectable SCR at input (dB)	Maximum phase step displacement (μm)
He-Ne laser	*632.8/0.002*	*20 cm*	*632.8/11*	1.05	-46.42	16.7 +/- 1.6
RCLED	*650/10*	22.77 μm	*651.9/36.2*	0.21	Less than -30	6.8 +/- 1.1
Tungsten halogen bulb with filter	*648.7/12.2*	11.48 μm	*651.9/36.2*	1.7	-31.96	2.9 +/- 2.0
LED	*644/18*	5.44 μm	*651.9/36.2*	2.5	-13.07	1.06 +/- 0.37

Table 5.1 - Summary of results from detection experiments, ordered by displacement

As would be expected in a coherence-based detection approach, there is a strong correlation between the coherence length of the source to be detected and both the minimum detectable SCR (MDSCR) and the maximum phase step displacement. This correlation, in the case of the MDSCR, is shown in fig. 5.12, where the coherence length axis is logarithmic to accommodate the large difference between the laser coherence length and that of the other sources. The RCLED MDSCR was not plotted because the figure in table 5.1 is not exact.

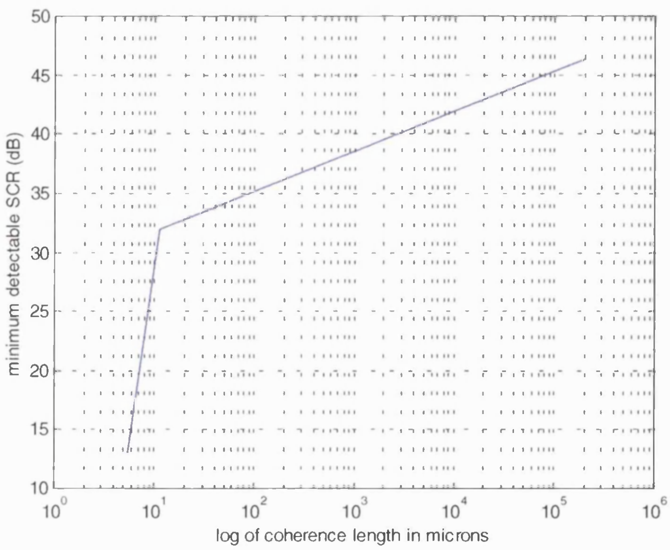


Figure 5.12 - Correlation between minimum detectable SCR and coherence length of the source

The correlation between minimum detectable SCR and coherence length of the source is explained in detail in chapter eight, where the theoretical model of the studied phenomenon is presented, and the curves given in figs. 5.8-5.11 are compared to theoretical predictions. The results obtained in these experiments show that IPSS displays very high sensitivities to coherent targets (e.g. gas lasers), and high sensitivities to partially coherent targets, when compared to other detection approaches reviewed in chapter two. In chapter seven, this comparison is made in detail.

5.1.3. Investigation of the optimum target-to-filter bandwidth ratio

As stated in chapter one, one of the aims of this research is the optimisation of the detection performance of the IPSS technique. In order to achieve this aim, a set of experiments was designed to investigate the influence of the optical filter bandwidth in the sensitivity to detect a target. This subject is also assessed theoretically in chapter eight. The experiments were carried out in the visible band, using red filters. The arrangement is shown in fig. 5.13. The tuneable target optical source consisted of a grating monochromator (Bentham M300), with a dispersion of 1.8 nm/mm, filtering the optical output of a 250 W white light source with a blackbody temperature of 3200 K. This gave a source with a variable central wavelength and bandwidth. The monochromator entrance and exit variable slits were set to 3 mm to simulate a target with a bandwidth of 5.4 nm. Sutton [Sut82] built a variable coherence length source using a monochromator, but kept the size of the exit slit constant, and varied the size of the entrance slit to obtain variable bandwidths. In the experiments described here both slits were set to the same width, and the coupling optics were adjusted to guarantee that the light from the sources filled the input slit. By doing this, it was possible to calculate the radiation optical bandwidth just by multiplying the slit width by the monochromator dispersion. A periscope was used to match the beam heights of the monochromator and the interferometer, and a plano-convex lens was used to collimate the beam. The background was simulated with a 20 W tungsten-halogen bulb having a similar blackbody temperature of 3220 K. As in the previous experiments, light from target and background was combined in a cube beamsplitter. The beam diameter was limited to 10 mm by an iris at the detection system's input. The detection system used was the same as in the previous sub-sections, with bandpass filtering in the audio amplifier between 100Hz and 1kHz and an oscilloscope sampling frequency of 10.24 kHz. Different SCRs were obtained by attenuating the radiation from the target using different power settings for the target 250 W light source, and attenuating the background using neutral density filters. The reason for this different arrangement is that, due to the low monochromator throughput, the target power at the interferometer entrance was very low. Hence, in order to obtain appropriate SCRs to observe the phase step shift effect, instead of placing neutral density filters in the target path, these had to be placed in the background path.

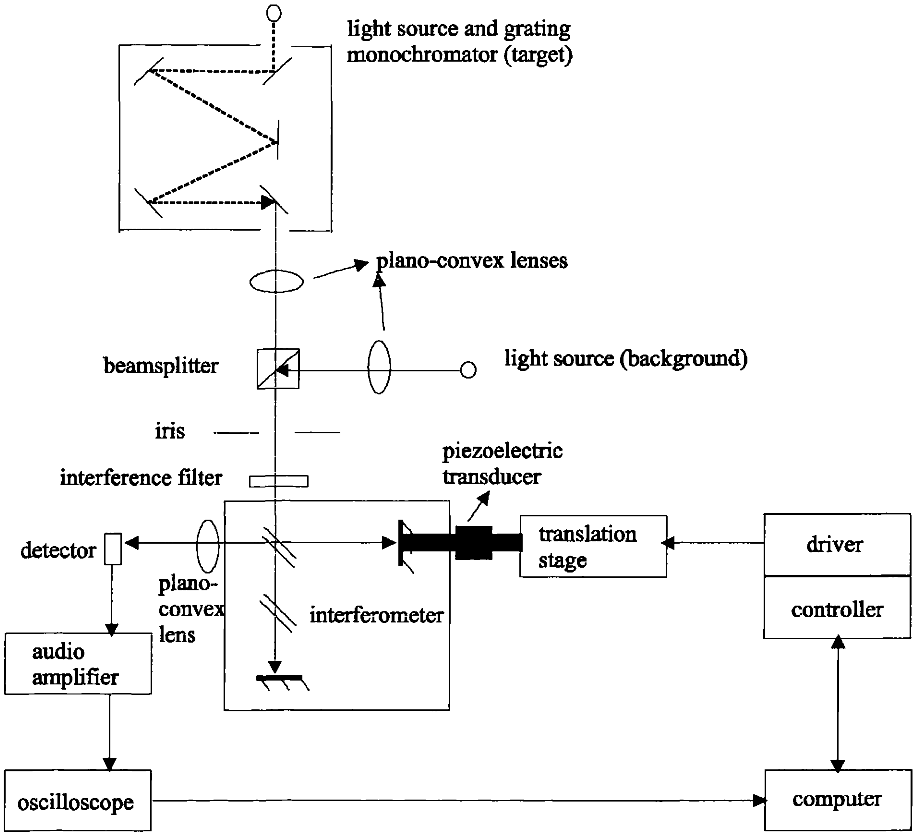


Figure 5.13 - Bandwidth ratio optimisation experimental arrangement.

In order to assess the effect of the target-to-filter bandwidth ratio, three target/filter combinations were used, which are shown in Table 5.2. The monochromator wavelength was varied in order to give targets with central wavelengths coincident with those of the available interference filters. As the coherence length is proportional to the wavelength squared (see eq. 3.3), this procedure slightly varied the coherence length of the target, but as the central wavelength variation between the three filters is small (+/-3%), this effect was considered negligible. In each target/filter combination, five different power ratios were tested and, as done in the previous sub-section, the five null feature displacements were combined together to give an average and a standard deviation per point.

Set	Central wavelength	Target bandwidth	Filter bandwidth	Ratio
1	632.6	5.4	11	0.491
2	651.9	5.4	36.2	0.149
3	674.8	5.4	17.8	0.303

Table 5.2. Target/filter combinations: all wavelengths and bandwidths in nm

Fig. 5.14 presents the phase step displacement as a function of the SCR for each of the three target/filter combinations. It can be seen that the maximum slope of the curve is obtained for a bandwidth ratio of 0.303, being smaller for ratios of 0.149 and 0.491, and indicating that there is an optimum ratio. This behaviour is confirmed theoretically in chapter eight, where these experimental results are compared with the theory.

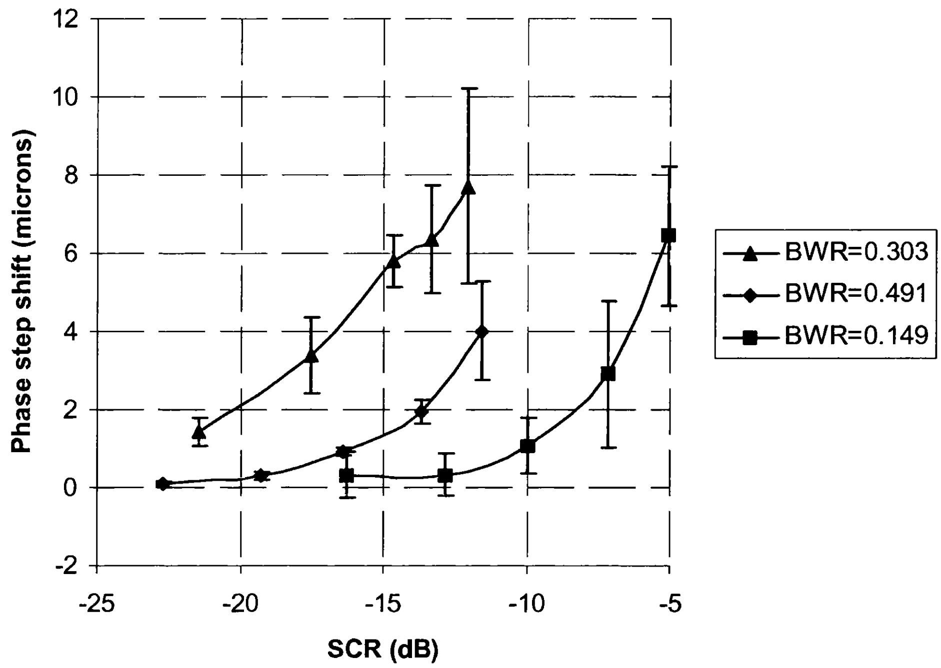


Figure 5.14 - Detection of a target with different filters. BWR is the target to filter bandwidth ratio.

In this bandwidth optimisation experiment, we have defined the detection responsivity as the slope of the power ratio vs. phase step shift curve, given in μm/dB. This slope was calculated using the shifts at the extreme points in the data, and the results are shown in fig. 5.15. The error bars showed in the plot where obtained by adding the variances of the shifts at these points, which make the error bars relatively large. In spite of this, the plot shows that the optimum bandwidth target to filter ratio is approximately 0.3. The conditions for which this optimisation is valid are presented in chapter eight.

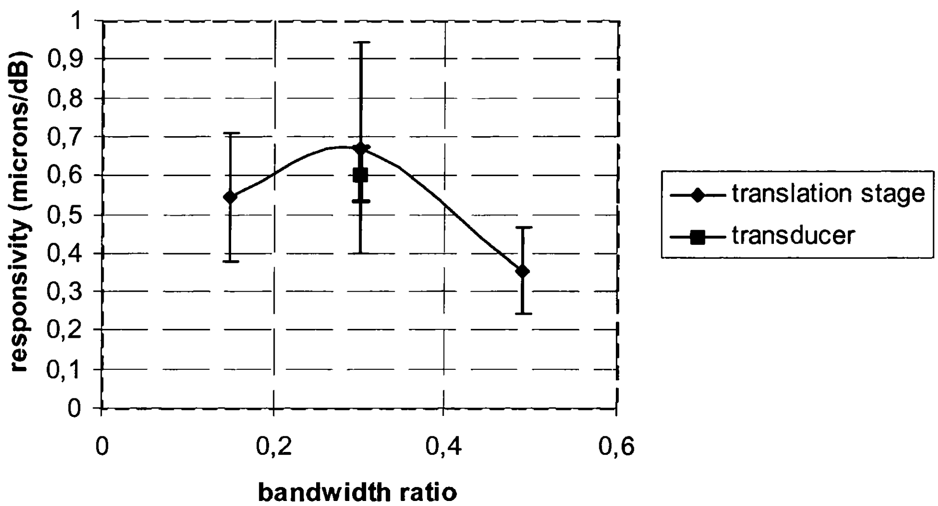


Figure 5.15 - Responsivity variation with bandwidth ratio

Due to the large error bar at the best bandwidth ratio (0.303), where one standard deviation was as much as 40% of the mean, the experiment was repeated using this ratio, near the peak, using a piezoelectric transducer for comparison purposes. To allow flexibility in locating the ideal path differences for the experiment, rather than replacing the translation stage with the transducer, the interferometer's moving mirror was mounted on the transducer that, in turn, was mounted on the translation stage. The mirror was scanned in a sinusoidal pattern at a frequency of 500 Hz, with fringes at 50 kHz. This high scan rate reduced the effect of intensity fluctuations in the light sources caused by the supply and of detector 1/f noise but required the use of a faster silicon detector. This experiment, which is described in detail in section 5.2, gave a differential responsivity of 0.603 μm/dB, against 0.671 μm/dB previously, but with a much-reduced standard deviation of 11.4% of the mean. Both measurements are in statistical agreement. The minimum detectable target-to-background power ratio or sensitivity, which was previously defined as the ratio which causes an average displacement just beyond one standard deviation, was −14.7 dB after filtering and −30.3 dB at the input, a very high sensitivity for a partially coherent target, comparable to the ones obtained with the filtered white light and RCLED measured in the previous sub-section.

5.1.4. Effects of target to filter wavelength offset

In the experiments previously presented, it was attempted to align the central wavelength of the detection system's interference filter with the target central wavelength. This was done by selecting adequate filters, and mounting them on a rotation stage, so that, by turning the filter, its central wavelength would change. The criterion employed in such alignment was to find a filter angle that would maximise the power transmitted by the filter. This does not guarantee that the central wavelengths are coincident. In order to study the limitations caused by an eventual target to filter wavelength offset, an experimental investigation was carried out to assess the effects of such an offset in the system responsivity. The filter in set no. 3 in table 5.2 was used, as it was closest to the peak responsivity in the earlier aligned wavelength experiment. Different target central wavelengths (665.9, 670.35, 674.8, 679.25 and 683.7 nm) were achieved by turning the monochromator grating. The results are shown in fig. 5.16. At offsets of 0.6% (4.5 nm in 674.8nm) the responsivity reduces to zero. Further offsets cause the displacement to become negative. This represents a reduction in coherence, as now the energy from the target is adding to the roll-off band of the filter, making the overall bandwidth larger, and the self coherence function smaller, as they are inversely proportional. The responsivity can be severely reduced if the prospective target central wavelength is not accurately centred in the optical filter spectral response, hence the penalty for achieving such high sensitivity is the need for previous knowledge of the target spectral characteristics. On the other hand, this change of sign in responsivity can be explored to create a combined OTIM/IPSS system, where a tuneable filter has its central wavelength continuously changed to create a modulation in the phase step shift, which could be used to obtain further sensitivity improvements. This study is suggested for future work. It remains to be established whether the optimum target to background bandwidth ratio remains optimum as the offset increases.

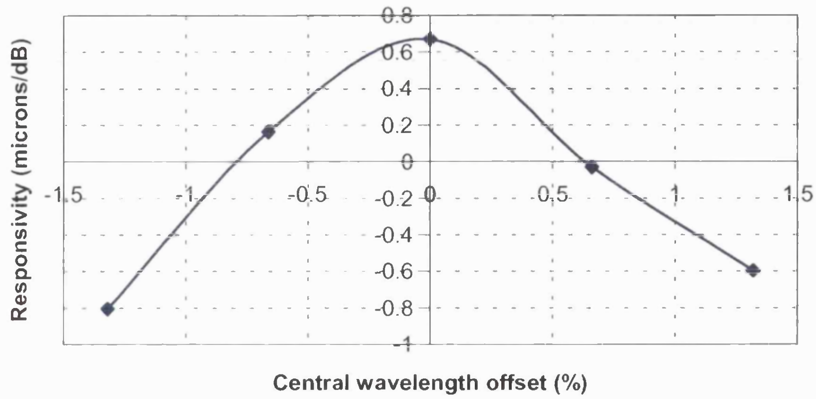


Figure 5.16 - Effect of target-to-filter central wavelength offset in detection responsivity

In order to make sure that the loss of responsivity was due to the offset, and not to the reduction of target power coupled through the optical filter, the target power was increased when the offset was maximum. The negative displacement of the phase step increased until the observability cut-off was reached, confirming that the loss was not due to a reduction in the filtered target power, otherwise the displacement had been positive.

Fig. 5.17 presents an interferogram recorded from a target with a central wavelength of 683.7 nm, and a filter centred at 674.8 nm. The plot shows a beating effect coming from the wavelength (and frequency) offset between target and filter, which is visible as an additional modulation on the fringe carrier. This effect is further discussed in chapter eight.

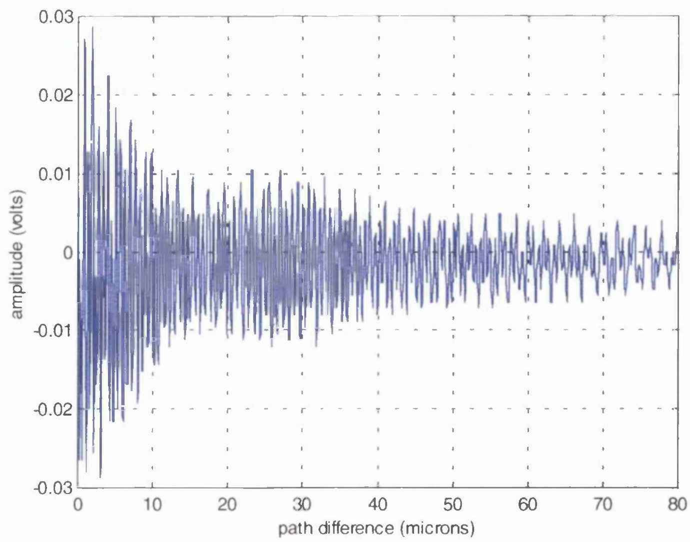


Figure 5.17 - Beating effect arising from target-filter wavelength offset

5.2. Experiments using piezoelectric transducers

5.2.1. Description, alignment and adjustments of a slow response transducer experimental system

As mentioned in the introduction to this chapter, different experimental arrangements were used to allow comparison, ensuring that the results obtained measure the phenomenon under study, and not experimental artefacts. The arrangement used in this section is similar to those described in the previous section, and is shown in fig. 5.18. The difference is in the path difference driving, or mirror scanning scheme. Instead of using a translation stage and a micropositioner, this arrangement used a piezoelectric transducer, controlled by a piezoelectric driver (Melles Griot 17 PCS 001). The transducer driving scheme does not suffer from the speed changes caused by the stepwise moving fashion of the translation stage, thus presenting less noise. The disadvantages are the lack of repeatability that occurs when a transducer without position feedback is employed, which is the case in our experiments, and the distortion caused by the non-linearities in the voltage vs. displacement characteristics of these devices.

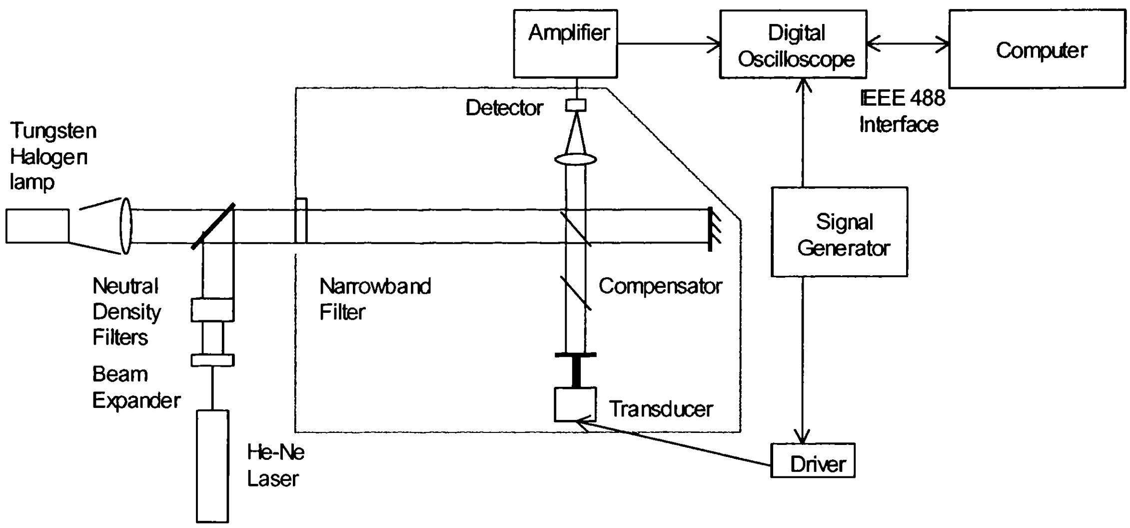


Figure 5.18 - Piezoelectric transducer experimental arrangement

As described in chapter three, one of the mirrors has to be scanned around the minimum visibility region of the interferogram. This is done by a translation stage attached to one of the mirrors. The stage is electronically moved by a piezoelectric transducer, whose high voltage driver output is reproducing a triangular waveform delivered by a signal generator. This signal is also sent to the oscilloscope for triggering purposes. The piezoelectric driver had a frequency response of 2 Hz (according to its manual), a figure checked by measuring the amplitude of white light interferograms as the scanning frequency was varied. Thus, the maximum mirror scanning frequency with this hardware was limited to 2 Hz, the reason to call this arrangement "slow response transducer".

To investigate which waveform would give the best scanning performance, experiments were done with four shapes: sinusoidal, triangular, sawtooth and sawtooth with idle time after flyback. The sinewave had the advantage of producing smoother interferograms, but suffered from fringe distortion caused by its intrinsic non-linearity. The sawtooth caused major distortions due to the high frequency components during flyback period; an idle time was introduced after flyback to allow the system to return to the initial condition, but as the time required to achieve that was up to one second, it turned out not to be a good choice. The triangular waveform shows the fringe pattern twice in a single scan, although with the hysteresis problem reported previously. The triangular waveform was then chosen, using the pattern from the rising edge and neglecting the one from the falling edge.

The term in eq. 4.1 which describes the change of length of the piezoelectric transducer with the voltage applied to it was measured by recording the voltage increase required to produce a certain number of laser fringes to pass through a reference point on a screen, and equals 0.488 μm/V. As the maximum voltage swing applicable to the transducer was 75 V, the maximum path difference scan obtainable with this arrangement was 0.488 μm/V x 75 V = 36.6 μm. This is further reduced by the beginning and end-of-sweep distortions, as happened with the translation stage in the previous section. It is not a limitation for IPSS, an approach that performs a narrow scanning around the minimum visibility section of an interferogram, but clearly a significant constraint to use the set-up to record an entire profile. This is the reason why, in chapter four, many short interferograms had to be spliced in order to record a long interferogram using a piezoelectric transducer.

The effects of averaging many interferograms using the digital oscilloscope are similar to the effects found in the experiments with a translation stage, described in the previous section. Figure 5.19 presents in the left side two sections (of interferograms) containing the phase step region. The path difference axis values are referred to the scan starting point, taken as zero. The top plot presents the non-averaged interferogram, with fringe amplitudes distorted by noise. In order to reduce the effects of noise and to allow the interferogram envelope to be seen at its minimum, the oscilloscope was set to average the data blocks, from ten to fifty times. The result made the minimum visible on the oscilloscope's screen, and considerably reduced the noise. The bottom left plot on figure 5.19 displays the interferogram after averaging, and the right side displays the corresponding spectra. These spectra were calculated using Matlab's " spectrum" function, which divides the data block into overlapping sections, removes the trends, applies 256 sample wide Hanning [Ope89] windows to each section, and completes it with zeroes. Then the sections are Fourier transformed through a discrete algorithm, and the squares of the magnitudes of the individual spectra are averaged to form the overall spectrum.

The effects of averaging were to make the minimum visible on the oscilloscope screen, which was useful to manually adjust the translation stage, and sample the correct interferogram region; to reduce the mean noise level, which dropped one order of magnitude (10 dB) from 500 to 2500 Hz (see right side of fig. 5.19); and to reduce the noise at frequencies close to the fringe's (signal) frequency, shown in the plot as a drop of 30 dB (10^3) at 250 Hz, decreasing with frequency up to 500 Hz.

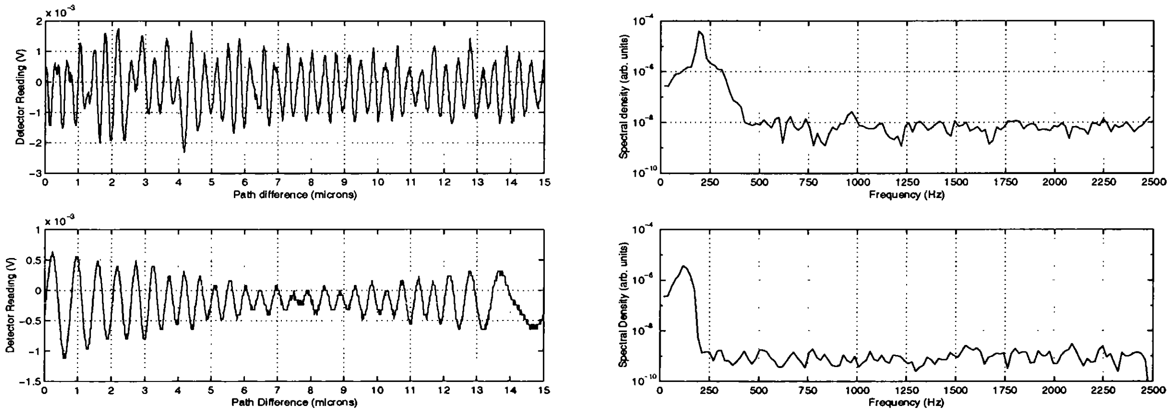


Figure 5.19 - Effects of averaging on the interferograms

The penalty of averaging was that the signal level was reduced by 10 dB at its peak, which aggravated the quantisation noise problem. This noise was more relevant in the region of interest, as the fringe visibility comes to a minimum there, and can be seen as small square oscillations in the lower left plot of fig. 5.19. To solve this problem, the level of signal delivered to the oscilloscope was increased through the use of an audio amplifier, so that it could occupy a larger portion of the oscilloscope's A/D converter's 10-bit dynamic range. This could be done either by increasing the level of light to be filtered, replacing the 20 W bulb by one with a higher power, or by introducing analogue amplification in between detector and oscilloscope. The latter alternative was chosen due to the following: the next available bulb power is 100 W, which generates a lot of heat that could influence the filter spectrum by changing its temperature; as its filament was much bigger, the 100W bulb represented a source with less spatial coherence, producing a less collimated beam, reducing fringe visibility at larger path differences; an increase in fluctuation noise due to the higher power could also be foreseen; and finally the fact that analogue amplification of the detector signal could be gain-controlled. The choice was to use an audio amplifier, with variable gain and a built-in analogue filter. This was another advantage, as the high frequency noise could also be removed before sampling and analogue to digital conversion, to avoid aliasing. The cut-off frequency for this filter was set to 1 kHz, and further digital filtering was required in order to recover the phase of the interferogram. Fig. 5.20 shows an interferogram recorded using a 40 dB voltage gain, low cut-off frequency of 1 Hz, and high cut-off frequency of 1 kHz. The region of interest now has much less squaring effects from quantisation noise.

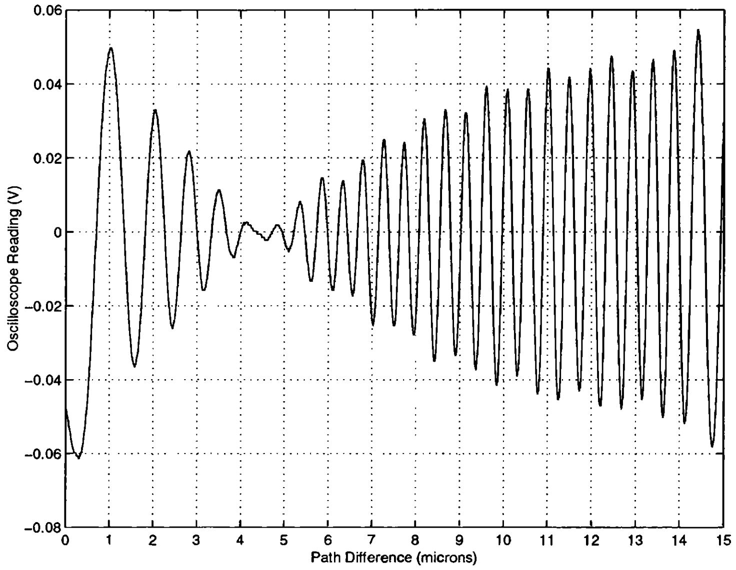


Figure 5.20 - Effects of analogue amplification and filtering

The interference filter was placed at the entrance of the detection system, and the position of the interferometer's moving mirror was manually adjusted, so that the interferogram's first minimum amplitude region was displayed on the oscilloscope screen. The ideal situation is that of figure 5.21, where this region is well in between the discontinuities caused by the reversal of the triangular driving waveform. In this figure, plot (a) is the triangular waveform produced by the signal generator, and plot (b) is the corresponding interferogram, acquired synchronously. The delay indicated by the arrows is due to the inertia of the mirror driving system.

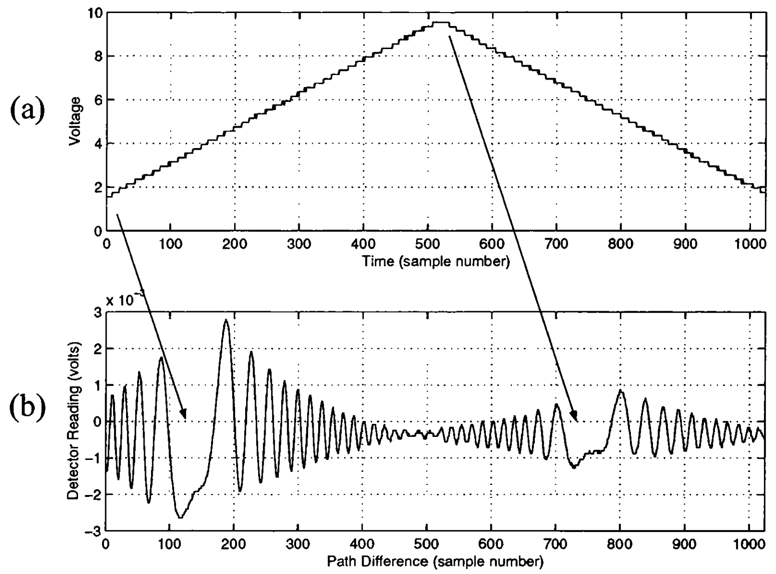


Figure 5.21 - Positioning of the phase step position within the mirror's scan. (a) - triangular waveform produced by the signal generator. (b) – interferogram.

Six combinations of neutral density filters with increasing attenuation were used to measure the phase step shift, as a function of SCR. Similarly to what happened in the experiments with a translation stage, the position of the phase step in the path difference axis varied even making successive acquisitions with a constant SCR, due to noise. Two components of this noise were identified: one systematic, represented by a constant drift in the shorter path difference direction; and one random, both caused by the piezoelectric transducer. The systematic error was that each time the stage came back to restart a scan, it came to a path difference a little shorter. The random component was due to transducer lack of repeatability and amplitude fluctuations in the light sources. The result of the systematic noise component can be seen in fig. 5.22, which shows the phase step, as extracted using the algorithm developed, for six different SCRs, the numbers on the ordinate axis being the laser target attenuation. As the laser power increases from top to bottom, the phase step was expected to move to the right, but as an effect of the systematic noise, it actually moved to the left (shorter path difference). Furthermore, for an attenuation of 38.4 dB, an example of a step due to (random) noise can be seen, with two phase steps very close to each other.

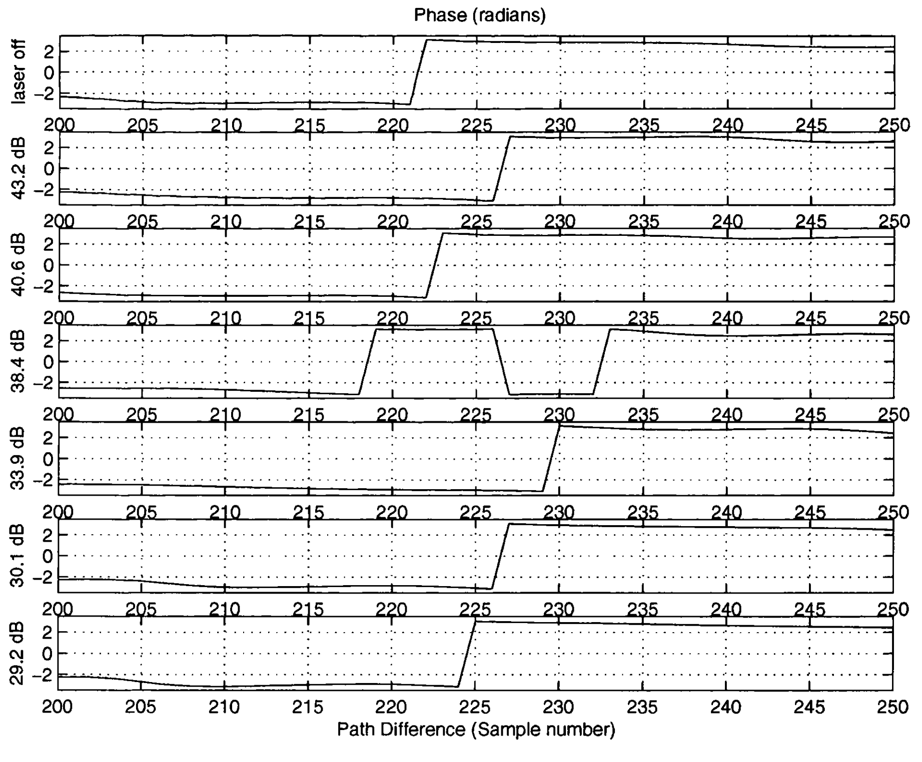


Figure 5.22 - Variation of phase step position as influenced by drift

To deepen the investigation of the drift in the phase step, four interferograms were acquired in five minute intervals, which are shown in fig. 5.23. The vertical line is crossing the phase step position in the lower plot, the first to be acquired, to show the drift to the left in the others. The three curves on the top were vertically displaced by adding a constant to all data samples.

Because of these effects, the acquisition of interferograms was done in pairs, as in previous experiments. Five pairs of waveforms for each laser attenuation were acquired, each pair containing one target-present and one target-absent waveform, both taken within 1 minute maximum time difference. This procedure minimised the drift problem, as both reference and signal were taken almost simultaneously. The IPSS algorithm was then applied to each pair, and the shift values were used to calculate an average and a standard deviation. These results are presented in the next sub-section.

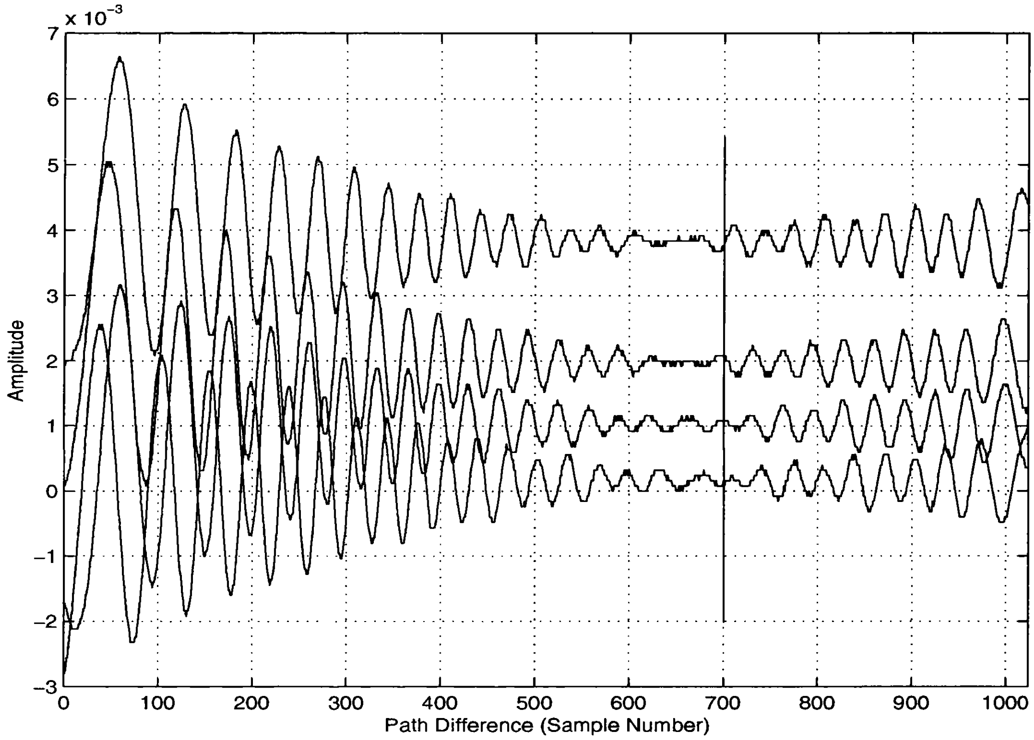


Figure 5.23 - Variation of scan start with time

Figure 5.24 presents one target absent/target present pair. In the three plots, the interferogram at the bottom is the target-absent waveform, while at the top the target is present, with an example SCR of -16 dB. Plot (a) presents the interferogram amplitudes, plot (b) the instantaneous phases, and plot (c) the instantaneous frequencies. In plot (c), the spikes were clearly discernible, and the position of the phase step could be detected by thresholding. However, other data blocks gave various spikes, or even worse, no spikes at all. Another interesting effect is the phase non-linearity in both waveforms, which may be a consequence of the averaging, as happened with the translation stage (see fig. 5.6).

(a)

(b)

(c)

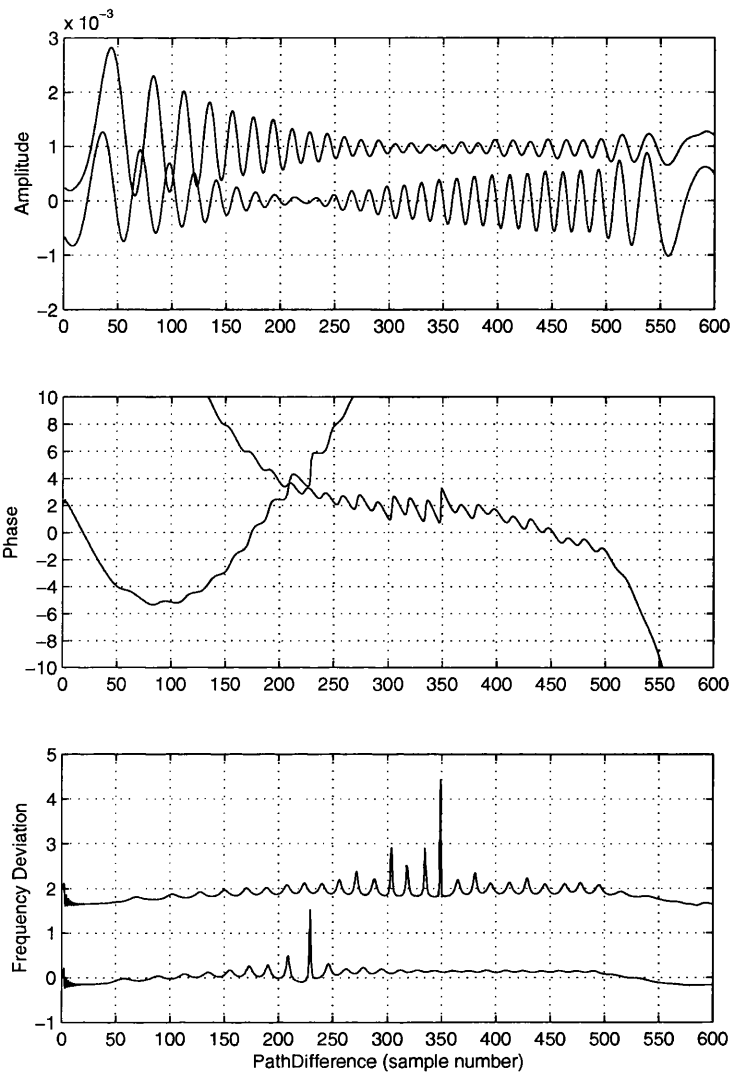


Figure 5.24 - Interferogram pairs showing the phase step shift, lower plots without the target. (a) – interferogram amplitudes; (b) – phases; (c) – instantaneous frequencies.

5.2.2. Measurements of phase step shift using the slow response transducer

Fig. 5.25 depicts the results of the application of the IPSS algorithm to the data acquired from a He-Ne laser using the slow piezoelectric transducer previously described. Up to –29 dB, there is no rising trend, and the error bars are large. From there, the coherence starts to rise as the signal is increased, allowing detection to be declared (using our simplified criterion) whenever the shift is higher than approximately 1 μm. In that situation the filtered SCR is –24.5 dB. The input SCR was not measured, but using an approximate figure of 18 dB for the SCR gain of this filter, taken from the power measurements done with the same laser in section 5.1, this SCR would be about –42.5 dB, again a very high sensitivity, but lower than the one obtained for the same target using the translation stage. The dispersion coefficient (standard deviation/average) was 1.29, about 23% more than the coefficient from the translation stage experiment, using the same light sources, showing the limitations of this mirror driving scheme. Another feature is that the maximum displacement of the phase step in this experiment was 6.4 μm, against 13.8 μm in the translation stage experiment (see fig. 5.8). This is due to the poor linearity and small path difference range of the piezoelectric transducer, which prevents longer interferograms (and hence, displacements) from being measured. In the conclusion section of this chapter, this difference is discussed in detail.

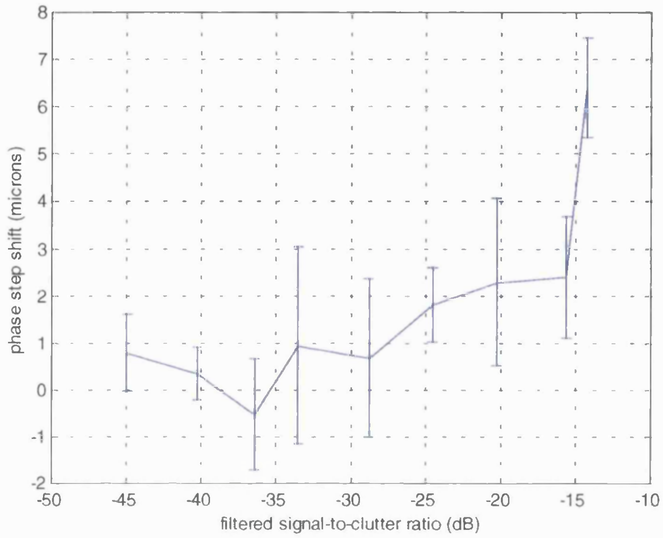


Figure 5.25 – Detection of a He-Ne laser with the slow piezoelectric transducer

Another experiment was carried out, to compare the detection system's performance with that of an intensity-based system when detecting the same target, in the same filtered background (and using the same power measurement system). The detector was placed at the input of the interferometer, close to the rear face of the interference filter. The tungsten halogen bulb simulated the background, and the same interference filter used in the IPSS experiment represented spectral filtering performed by the intensity system. The power meter maximum and minimum readings in a period of 1 minute were 24.70 and 24.45 μW (-16.07 and –16.12 dBm respectively), simulating the various noise processes involved in the amplitude detection method. The target was then attenuated until the power meter gave a reading just above noise fluctuation (24.80 μW or –16.06 dBm), simulating the minimum detectable signal. This corresponded to a configuration of neutral density filters giving 29.31 dB of attenuation. The power measured from the laser, through the interference filter was –4.72 dBm, giving a signal power of –34.03 dBm at the interferometer input. Taking the average background power as –16.10 dBm, the sensitivity of the amplitude method was –17.93 dB. Thus, the IPSS method had a sensitivity 11 dB better than a conventional, intensity-based method relying solely on a power meter's internal signal processing.

The experiment in this sub-section demonstrated the feasibility of using piezoelectric transducers in IPSS with relatively good sensitivity. However, the transducer's repeatability and speed severely limited the performance of this arrangement. The lack of repeatability produced unwanted displacements of the phase step not due to target presence, and the low speed (1.5-2 Hz) produced low fringe carrier frequencies (approximately 150 Hz), which are in the region of strong detector 1/f noise and of noise from the light bulb's supply flicker (see fig. 5.19). An additional inconvenience of the low scanning frequency is the low update rate, or the low number of times per second a new interferogram is formed. This is mainly important in real time applications.

5.2.3. Description, alignment and adjustments of the fast response transducer experimental system

Both the translation stage presented in section 5.1, and the piezoelectric transducer shown in 5.2.1 present mirror scanning frequencies around 2 Hz, and fringe carrier frequencies around 150 Hz, with the problems described previously. There was, then, a need to investigate the IPSS technique with faster mirror scanning, to obtain higher fringe frequencies, away from these disturbances. In this sub-section a much faster transducer is employed, capable of achieving scanning frequencies 1000 times higher, and is called "fast response" transducer. In these experiments, an arrangement (fig. 5.26) similar to the one used in section 5.1 was employed, in which the transducer's vibrating rod was screwed onto the mirror mounting, while its main body was screwed to a mounting attached to the translation stage used in section 5.1.

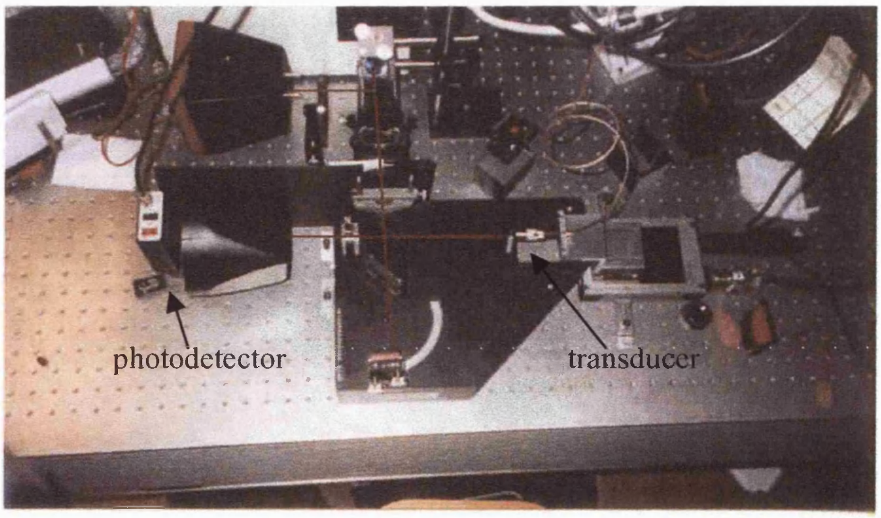


Figure 5.26 - Fast piezoelectric transducer arrangement. Red lines indicate light paths.

This arrangement allowed the use of the translation stage for large path difference bias, and the transducer for the AC mirror driving signal. A Physik Instrumente P 842.10 transducer was driven by an E-505.00 LVPZT (low voltage lead zirconate titanate) amplifier module of the same make. Although still an open loop transducer, the P 842.10 displayed a much better performance than the previous transducer. It has a 15 μm travel, enough for IPSS, 0.3 nm resolution and an unloaded resonant frequency of 18 kHz, allowing sub-millisecond response times. Another modification with respect to the previous arrangements was the use of a fast response Silicon PIN photodetector (New Focus 2031, shown in fig. 5.26), as the detector used in previous experiments did not have the bandwidth required. The device has an active diameter of 8 mm and three different gains, with bandwidths of 1 MHz, 150 kHz and 90 kHz, the latter corresponding to the highest gain, keeping the gain-bandwidth product constant. This internal amplification eliminated the need for an external amplifier. Once higher carrier frequencies were achieved, the dominant detector noise changed from 1/f to shot noise caused by stray light. We have checked this by using a sheltering cap with an adequate aperture (seen in fig. 5.26) in front of the detector to reduce the amount of stray light reaching it, which indeed reduced the noise in the oscilloscope trace, with the light sources switched off.

As in previous experiments, the interferometer's mirror was adjusted to give circular fringes that would produce maximum modulation efficiency. However, the pressure of the transducer's tip on the mirror mounting caused a deformation in this mirror, and the fringe format, depicted in fig. 5.27, can be seen to display such deformation, as if the mirror was changed into a plano-convex lens. The image was acquired by a CCD camera placed at the interferometer's exit plane, connected to a frame grabber, both used in the experiments in chapter six.

Figure 5.27 - Fringe format in the fast piezoelectric transducer experiment.

As in the previous sub-section, a set of experiments was performed to decide on the best waveform to be used. Due to the very high scanning frequencies employed, the use of triangular waveforms would create harmonics that would reach the transducer's resonant frequency, with risk of damaging it. Hence, in this case, the choice was the use of sinusoidal driving signals, in spite of the corresponding loss of linearity. More experiments were performed to choose an appropriate scanning frequency. Although the unloaded resonant frequency was 18 kHz, measured with the transducer's effective mass of 10 g, this frequency was reduced to 4 kHz with the addition of the mass of the mirror and its mounting, which increased the weight to 36 g. Another limitation to the maximum scan frequency came from the maximum current supplied by the LVPZT amplifier, which posed a trade-off between transducer displacement and scan frequency. For example, with the transducer's maximum displacement of 15 µm, the maximum frequency would be 1.33 kHz. By reducing the transducer's excursion to 10 µm, the maximum frequency is increased to 2 kHz. In order to have a larger dynamic range in the IPSS technique, it was decided to use the maximum excursion, 15 µm, with a scanning frequency of 500 Hz. These settings allowed the experiments to have fringe frequencies of about 48 kHz, well above detector noise, but without pushing the transducer to its endurance limits.

As in previous experiments, the optics were aligned in order to direct the beams from target and background to the active surface of the photodetector, through the aperture of the shielding cap used to reduce background noise. The photodetector gain was set to medium, providing enough signal gain, with an adequate bandwidth (150 kHz). Sample interferograms were then recorded and the displacements in the phase step with the presence of a laser target observed.

The next investigation concerned transducer repeatability. Fig. 5.28 presents interferograms acquired with 1 minute between acquisitions, from top to bottom. Using the IPSS algorithm, there was an average drift in the phase step path difference of 27 samples/minute. This drift is due to the open loop operation of the piezoelectric transducer, that is, without position feedback. It required the data to be acquired in target-present/ target- absent pairs, as done previously. As the target-present interferograms were acquired after the target-absent ones, the drift tends to reduce the measured phase step displacements. Using the sample number to microns conversion factor of 0.0138 µm/sample, calculated using eq. 4.1, this drift corresponds to 0.3726 µm/minute. The experimental data was acquired with a maximum delay of 15 sec, making the maximum error 0.093 µm, negligible when compared to the system's random errors.

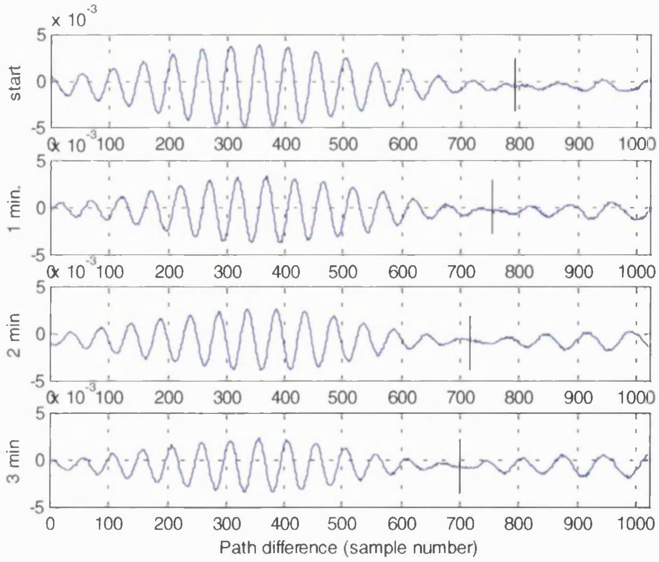


Figure 5.28 - Fast piezoelectric transducer drift effects . Vertical lines represent estimated phase step points.

5.2.4. Measurements of phase step shift using the fast response transducer

Two targets were employed in the detection experiments with the fast piezoelectric transducer: a He-Ne laser and a monochromator, both used previously. The measurements were taken using the same procedure: five interferogram pairs per SCR, calculating the average and standard deviation of the phase step displacement for each of the thirteen SCRs. Figure 5.29 presents the results. Up to SCR = –6 dB, the curve has the same shape as the ones obtained using previous methods, although with alternating signs of the second derivative. This behaviour is believed to be due to the use of a sinusoidal mirror motion. The expected behaviour would be one measured with a triangular waveform, which has a linear ramp characteristic. The problem can then be seen as the linear approximation of a sinusoid. In its central portion, the sinusoid is almost linear; as one departs from the centre, the approximation worsens and distortion happens. Longer path difference displacements appear to be shorter, forming the "bend down" effect on the plot. Due to the drift observed and described previously, different delay times in the oscilloscope were used in order to keep the region of interest (where the phase step occurred) within its acquisition window. Because of this, different portions of the sinusoidal characteristic were then sampled, causing different degrees of distortion throughout the curve, preventing the use of a calibration procedure to remove this distortion. Hence, the sinusoidal waveform is not a good choice for the purpose of comparing experimental results to theory. It has been used in its case solely to allow the use of high frequency fringe carriers, away from noise.

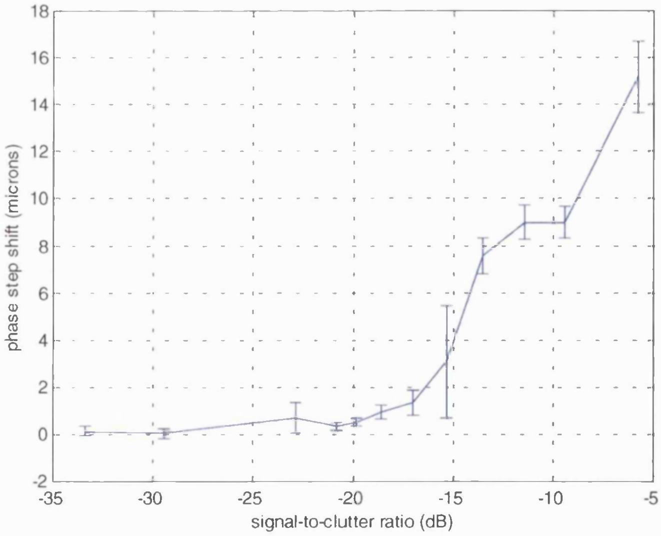


Figure 5.29 - Detection of a He-Ne laser using the fast piezoelectric transducer

The sensitivity in this case was –22.86 dB, corresponding to –40.86 dB at the system's input, a reasonably good figure, although not the best one. The average dispersion coefficient was 0.85, the lowest of all laser measurements in this chapter. Thus, the fast piezoelectric transducer, although not being the most sensitive method, was the one with least noise, as would be expected, as it eliminates the noise associated with low frequency operation (1/f and 100 Hz supply flicker). This fairly good performance was achieved with an update rate of 500 times/second, demonstrating the suitability of the technique to real-time applications. It is believed that the fast transducer arrangement was not the most sensitive because of the transducer non-linearity, which may have caused a slight increase in the error bars at smaller SCRs than the one considered.

Another target detected using the fast transducer was the monochromator, in the most sensitive configuration studied in 5.1.3. It acted as a filter with a central wavelength of 674.8 nm, with a bandwidth of 5.4 nm. The detection system's filter had the same central wavelength, and a bandwidth of 17.8 nm. Due to the low throughput of the monochromator, it was necessary to introduce further amplification apart from the photodetector's internal gain. This need was met by the same amplifier employed in 5.1, with a gain of 30 dB, and cut-off frequencies of 1 kHz and 100 kHz. Additionally, eight interferograms were averaged in the oscilloscope to further increase the SNR. Six SCRs were measured. The results are depicted in fig. 5.30, where only a slight curve distortion (bend down) is seen. The sensitivity was -14.7 dB, corresponding to an input SCR of -30.24 dB, a figure almost identical to the one obtained in section 5.2. The dispersion coefficient was 1.5, higher than in the laser case, due to the low monochromator throughput and to noise added by the amplifier.

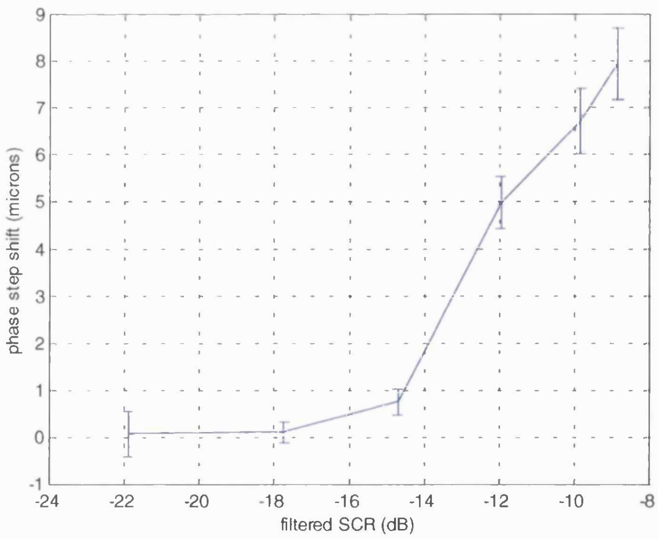


Figure 5.30 - Detection of a monochromator target using the fast piezoelectric transducer

5.3 Conclusions

This chapter presented the measurements of phase step displacements in temporal interferograms, generated by longitudinal movement of one interferometer mirror. Three experimental arrangements were studied: one with a micropositioner implemented with a motorised translation stage, and two using piezoelectric transducers, of which one was slow, with a scanning frequency of approximately 2 Hz, and the other scanning the mirror at approximately 500 Hz. Fig. 5.31 presents a comparison between the results obtained from the same light source (He-Ne laser), using the three arrangements. The error bars are not displayed to make the plot clearer. Up to an SCR of −20 dB, the fast transducer and translation stage do not depart from zero, and stay in good agreement up to −15 dB, from where the agreement is only modest, due to the non-linearities in the transducer. The slow and fast transducer results coincide around − 15 dB, where the curve slope changes very rapidly. Below this range the slow transducer curve presents an irregular shape due to its higher noise level. The slow transducer curve ends at −14 dB because at the time the experiment was performed (see [Cou99]) the only interest was to assess the minimum detectable SCR, and no points were taken at larger SCRs. The observability cut-off only happens in the other two curves, between −10 and −5 dB. The good agreement between the experimental results obtained with three different arrangements (a fourth one is included in the comparison in chapter six) demonstrates that these results are reliably characterising the phenomenon under study.

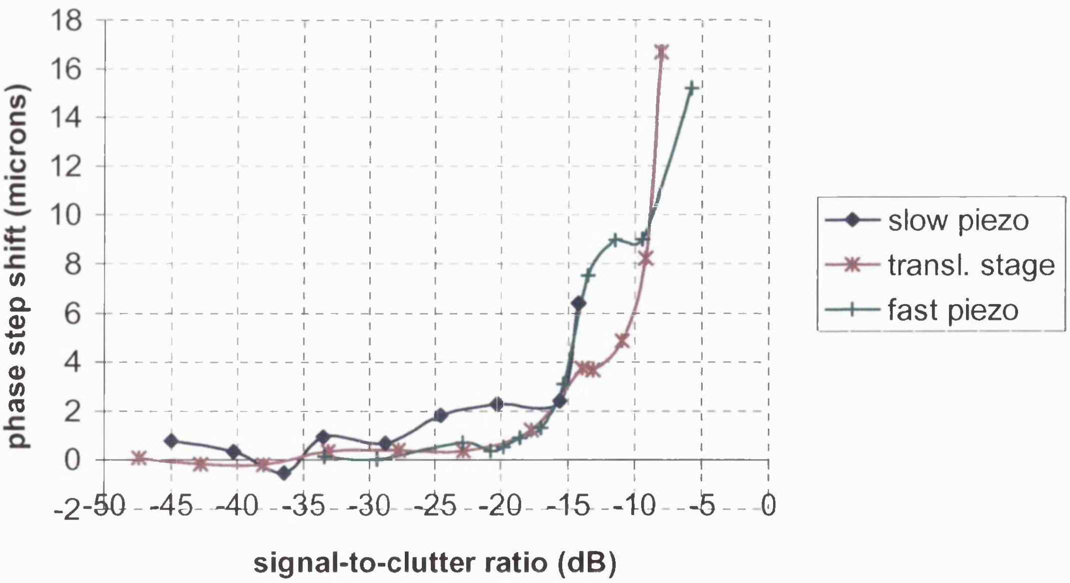


Figure 5.31 - Comparison of laser results using different arrangements

The arrangement that displayed the highest sensitivity was the translation stage one, the transducer arrangements suffering slight degradation due to noise and non-linearity (slow arrangement) and non-linearity only (fast arrangement). The fast transducer had a very small sensitivity loss when compared to the slow transducer, thus exchanging a much higher update rate for a graceful degradation in sensitivity. This arrangement demonstrated its suitability for real-time applications, with a good balance between speed and sensitivity. Interestingly, this approach had the smallest dispersion coefficient, due to its avoidance of the 1/f noise region. The sensitivity increases with target coherence length in all arrangements, as expected from a coherence-based approach.

A study of the influence of the target-to-filter bandwidth ratio was performed, and demonstrated that there is an optimum ratio, at approximately 0.3, using the criterion of the largest responsivity of the phase step shift with relation to an increase in SCR. A similar study demonstrated that the effects of target-to-filter central wavelength offset are the reduction of responsivity, and even the inversion of the direction of displacement, which can be exploited for further improvement of the approach.

While the piezoelectric transducer solution appears to be the most adequate for high update rates, real-time applications, the best experimental approach to allow comparison with theoretical results is the translation stage arrangement, due to its better linearity characteristics. The comparison between the phase step displacements, optimum bandwidth ratio and effects of central wavelength offset, measured with the translation stage, and the theory developed in chapter eight, is also presented in chapter eight.

Chapter 6

Imaging Detection Experiments

In chapter five, the high sensitivity of the Interferogram Phase Step Shift (IPSS) approach was demonstrated through experiments where the presence of coherent or partially coherent targets was declared from interferograms with one single, temporal dimension. In order to detect a target, the instrument would have to be pointed towards it. It is also possible to think of the non-imaging experiments presented in chapter five as an assessment of the per-pixel sensitivity of an imaging system. This chapter presents imaging detection experiments arranged in two different embodiments, both employing a CCD camera replacing the photodetector used previously. The first operates on images of interference fringes, formed on the CCD array, to extract one-dimensional interferograms. These signals are then processed as was done in chapter five to declare detection, and the results are compared to those obtained there. The second experimental set comprises a coherence imaging system implemented to demonstrate that the very high sensitivity of the IPSS technique is achievable even in imaging applications. This experiment applies the IPSS approach to every pixel in an image, forming another image from the coherence contrast between target and background. In the second experiment, the minimum detectable SCR (MDSCR) is found using a visual detection criterion. As some of the work done previously in the area of coherence-based detection did not quote detection sensitivities [Sut82] [Hic88], the chapter includes the reproduction of experiments done in earlier work, whose results are used for comparison with IPSS sensitivity. It is demonstrated that IPSS has better performance than existing coherence-based and amplitude-based approaches. It is believed that the results presented in this chapter are the first effective demonstration of imaging based on target-to-background coherence contrast.

6.1. Description, alignment and adjustments of the experimental system

The experimental arrangement used in this chapter is presented in fig. 6.1, and introduces slight modifications when compared to the arrangement used in section 5.1 (translation stage experiments). A zoom lens (Computar 501 4377) was placed at the exit of the Michelson interferometer, its field of view being adjusted according to the requirements of the different experiments, such as magnification and focal distance. The lens images the interference pattern formed in the interferometer's exit plane into the array of a monochrome CCD (charge-coupled device) camera (Sony XC-77RR-CE), which outputs an interlaced CCIR analogue video signal to a frame grabber (Matrox Pulsar), installed in a computer. The same computer also controls the interferometer's moving mirror using the same scheme used in section 5.1, comprising a motor controller and a translation stage. Automatic gain control (AGC) and gamma correction, although available in the camera, were not used, to avoid non-linearities that could affect the results of the experiments.

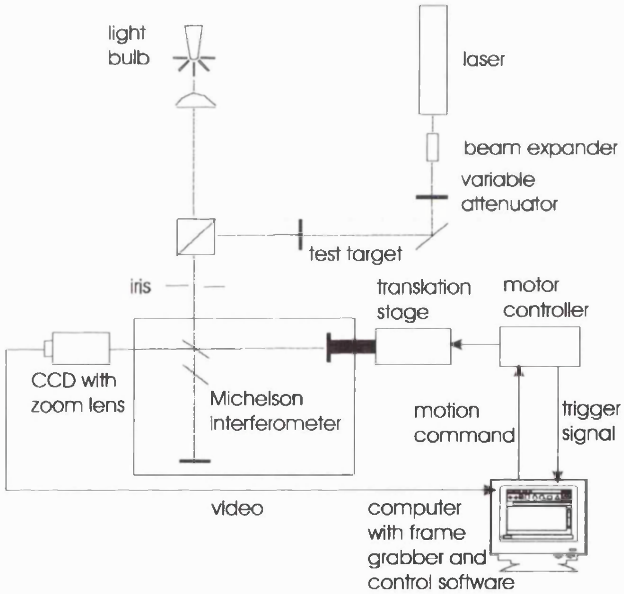


Figure 6.1 - Coherence-based imaging system experimental arrangement. The dotted lines indicate the light path. An interference filter is included between the iris and the interferometer in some experiments.

In one of the two operation modes studied in this chapter, a horizontally increasing path difference is created in the CCD array by tilting one of the interferometer's mirrors about a vertical axis passing through one of its edges. The resulting interference pattern is known as Fizeau or line fringes [Hec98]. This technique has been used in Static Fourier Transform Spectrometers (SFTS), described in chapter 2, displaying the advantage of using no moving parts [Iva00]. To record the spatial interferograms to be processed by the IPSS technique (in the next section), the interferometer's tilting mirror was adjusted to give vertical, line fringes on the CCD array, and the focal length of the zoom lens (adjustable from 18 to 108 mm) was set to produce fringes filling the whole array surface, as shown in fig. 6.2. The entrance iris aperture diameter was 17 mm, which allowed imaging onto the whole array without losing light out of it. Fig. 6.3 depicts an image of horizontal fringes produced in the alignment phase using a 20 W tungsten halogen lamp simulating the background, filtered by an interference filter (not shown in fig. 6.1) with a FWHM of 36 nm and a central wavelength of 651.9 nm. The path difference, which in fig. 6.3 increases from bottom to top, was adjusted so that the acquired image included the region of interest in the interferogram, where a fringe visibility minimum occurs. Looking at fig. 6.3 from bottom to top, it can be seen that the visibility increases, goes through a maximum in the ZPD (zero path difference) region, and then starts decreasing. An arrow indicates the region where fringe contrast goes through a minimum. In this area, a careful observation shows a change in the period of the fringes, which corresponds to the phase step explored in IPSS. Further to the top of the image, the visibility increases again, and then finally decreases, representing the first sidelobe in the self-coherence function. If the greyscale values of the pixels in a column of the image are measured, the resulting vector will represent an interferogram whose path difference was created spatially, rather than temporally as done in chapter five. To allow a comparison, a plot of the self-coherence function of this filter (central wavelength 651.9 nm and FWHM 36 nm), generated temporally, was given in chapter four, fig. 4.10.

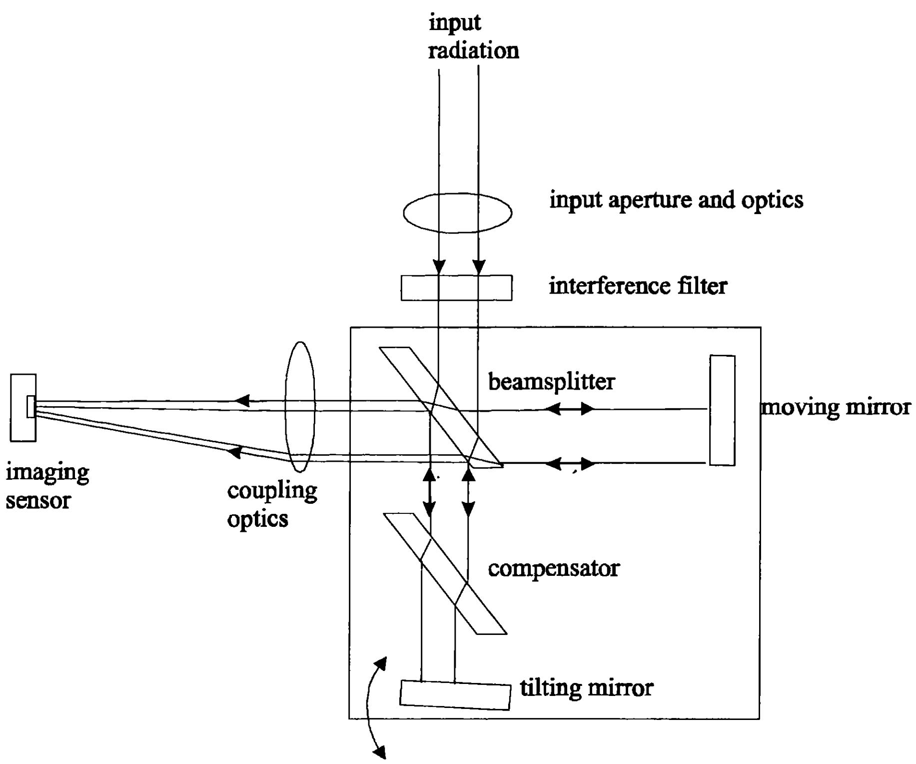


Figure 6.2 - Optical beam ray tracing in the CCD arrangement

In order to avoid saturation in the CCD, the linearity between the light levels coming from the lamp, and the corresponding voltages in the CCD array was assessed, and the maximum non-saturating lamp power level established. Light powers from the targets used in the experiments were a lot lower, hence of little concern to saturation. Two types of target were used in this chapter: a 5 mW He-Ne laser and a grating monochromator (not shown in fig. 6.1), both already used in chapter five. In fig. 6.4, the fringes from the laser can be seen, with speckle coming from the reflection on the surfaces of the optical system and from residual dust representing a considerable source of noise. Other important noise sources in these imaging experiments, affecting both types of target, were background/fixed pattern noise, and the CCD array's own noise. The fixed pattern noise was caused by illumination spatial non-uniformities, bad or dead pixels in the CCD array, and dirt on the optics, which was present in spite of regular optics cleaning. In order to reduce the effects of fixed pattern noise, reference images without fringes were acquired and subtracted pixel by pixel from the images containing fringes, both images being acquired using the same illumination and optics. Instead of pixel by pixel division, which would have been used if the aim was to display the best possible image, subtraction was employed to separate the AC signals containing the interferogram information from the high intensity illumination background, dead pixels and dirt seen in fig. 6.3, hence generating interferograms similar to the ones used in chapter five.

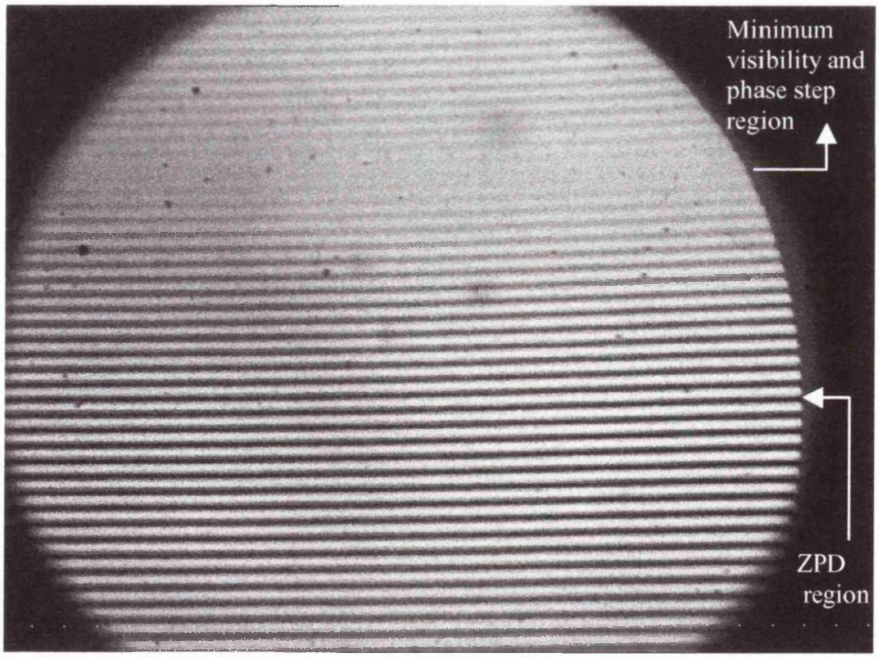


Figure 6.3 - Interference fringes from a 20 W tungsten-halogen lamp, filtered by an interference filter with a FWHM of 36 nm, central wavelength 651.9 nm

The reference images were obtained by two different methods: in the images without a target and with a monochromator target, the path difference was increased to a value much longer than the source's coherence length, where no interference occurred. An image was then acquired, ensuring that the same illumination conditions between reference and signal images were met. The subtraction operation corresponds to the removal of the first two terms in the General Law of Interference (eq. 3.1), leaving the interference term, which is the signal of interest. When the target to be detected was a laser, the method just described could not be employed, as the interferometer's maximum path difference, twice the translation stage's maximum excursion (2 mm), was much smaller than the target's coherence length (~ 20 cm). The method of choice was, then, to block the tilting mirror of the interferometer, acquire an image of the reflection from the moving mirror (which was not moving in this case), and to apply a correction factor of 1/0.46=2.17 instead of 2 due to the beamsplitter's uneven splitting ratio, presented in chapter four (46%-54%). This method caused some degradation in the extraction of the interferogram signal, as will be detailed later. After the removal of the fixed pattern noise, the CCD's dark current shot noise was increased by the combination of reference and signal image variances, thus becoming dominant. This noise was reduced by the software filtering described in the next section.

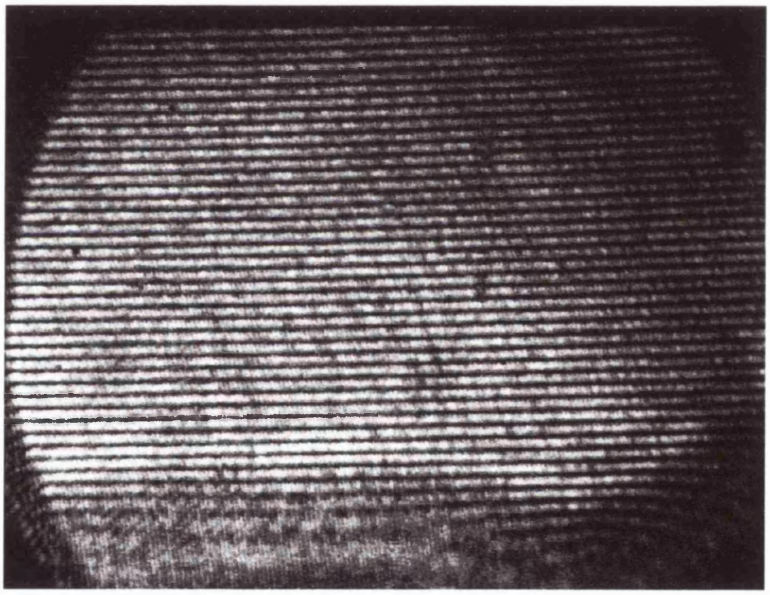


Figure 6.4 - Fringes from the 633 nm spectral line of a He-Ne laser

Classical image processing techniques, such as maximum or median filters, edge enhancement or histogram equalisation [Wee98] were not employed in this research. Although they are of importance in applications where an image has to be presented to a human observer with the highest quality possible, most such techniques are non-linear, and their use would prohibit comparison with non-imaging IPSS experiments. Nevertheless, Wiener filtering [Wee98] was employed after background subtraction, to improve the signal-to-noise ratio. Assuming the noise in the image is additive and Gaussian, which is reasonable as the dark current shot noise does not depend on signal power [Der96], this filter estimates its mean and standard deviation in the neighbourhood of each pixel, and uses these estimates to low pass filter the image, removing the estimated noise. Fig. 6.5 shows the image processing steps used prior to the application of the IPSS algorithm. In the image a) a dim spatial structure can be seen on top of the background, which is shown in b). In image c), the interference fringes are more visible, but noise is larger than in a), due to the addition of the signal and reference noise variances. Finally, in d) a Wiener filter with a neighbourhood of size of 5 by 3 pixels was applied, reducing but not eliminating the noise. This filter window size was selected as follows: if it was too narrow, the estimation of the noise parameters would be poor due to insufficient data; if the window was too large, the same would happen, but now because the filter would combine the (different) noise statistics of dark and bright fringes. The selected size, then, was aligned to the fringe orientation.

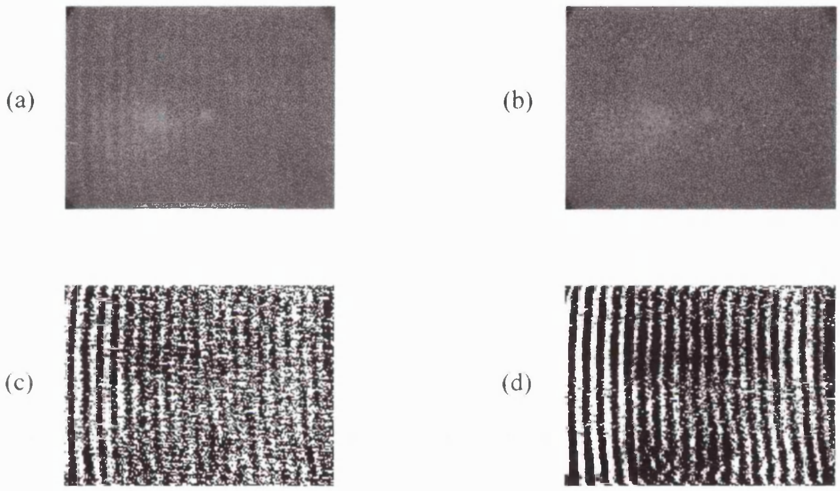


Figure 6.5 - Image processing operations applied to the images: a) recorded interferogram (signal); b) reference image; c) signal minus reference; d) Wiener filtering

In the experiments done in chapter five, investigations of the drift caused by the translation stage and piezoelectric transducers were done. In the experiment generating spatial interferograms, however, there are no repetitive scans, and hence, no reason for such drifts to happen. To certify that this was true, a sequence of five images containing fringes from the tungsten halogen bulb, filtered as described previously, was taken with one-minute intervals. The drifts in the fringe profile were measured using the IPSS technique, and no continuous drift to any side was found. The absence of such drift is another advantage of the spatial approach, when compared to temporal: better stability of the fringe pattern with respect to path difference, a parameter that is critical for delay-measuring systems such as IPSS.

6.2. Application of the IPSS technique to spatial interferograms

As mentioned in the previous section, two light sources were used as targets in the experiments applying the IPSS technique to spatial interferograms: a grating monochromator representing a partially coherent target and a He-Ne laser, a coherent target. The images recorded with the CCD camera had 768 horizontal and 576 vertical lines. In order to obtain the highest spatial sampling frequency possible, vertical line fringes were generated at the interferometer, by tilting one of its mirrors about a vertical axis. This procedure allowed the exploration of up to 576 interferograms, each one with 768 samples. Fig. 6.6 presents the fringes obtained from the Bentham M300 monochromator, illuminated by a 250 W white light source with a blackbody temperature of 3200 K. In the experimental arrangement seen in fig. 6.1, the laser and beam expander are replaced by the monochromator, a periscope to match the monochromator and interferometer beam heights, and collimating optics. Due to imperfections in the optical components or in their alignment, the fringes are not perfectly straight, with the strongest distortion close to the top and bottom of the image.

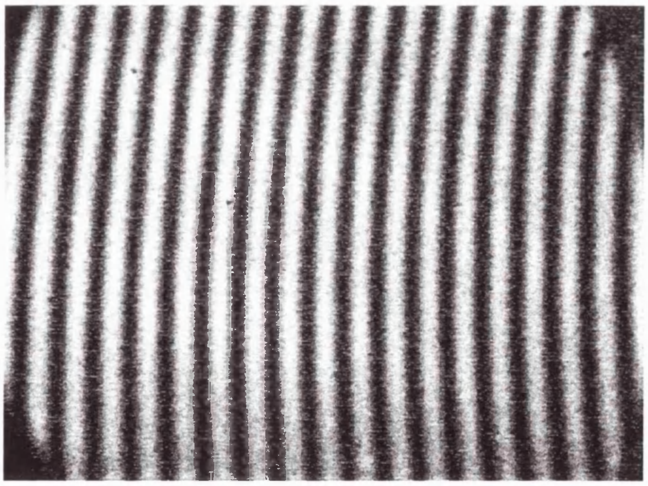


Figure 6.6 - Fringes from a monochromator with a central wavelength of 674.8 nm and a FWHM of 5.4 nm. This image was enhanced using histogram equalisation, for visualisation purposes only.

The algorithm used in this set of experiments comprised two tasks: the image processing module extracts one-dimensional interferograms from the images, and the IPSS module performs target detection using these interferograms, and the same algorithm employed in chapter five, and described in chapter seven. With both targets, five image pairs were acquired per SCR, each pair containing an image without the target, and another with the target present. With the monochromator target, six SCRs were measured. The image processing module loads a fringe image pair and one reference image, acquired as described previously, to remove fixed pattern noise. As the target is much dimmer than the background, a reference image with the target present is not required. The reference image, acquired with the background only, is subtracted pixel by pixel from both fringe images, and a Wiener filter is applied to the resulting images. From each image, one-dimensional horizontal profiles are formed by taking the intensity of the pixels in a horizontal line, for a number of lines in the centre of the images, where the distortion is reduced. These profiles, with 768 samples each, are then averaged to produce an interferogram, to be processed by the IPSS module. Fig. 6.7 shows a sixty-line strip from a fringe image acquired from filtered light radiated by a tungsten halogen bulb, higher brightness shown in white. The image clearly shows a phase change indicated by the arrow, at the same path difference where a spike in instantaneous frequency is seen in the bottom plot.

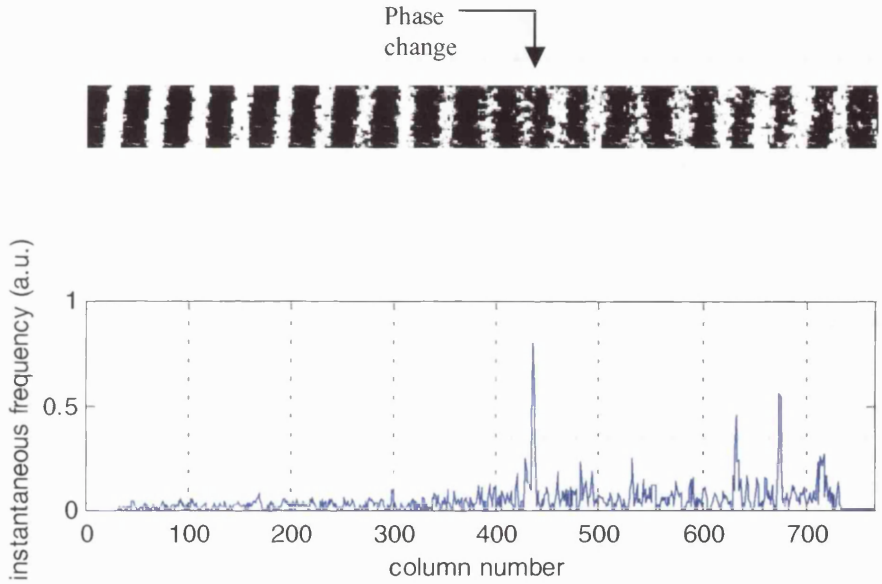


Figure 6.7 - Above, an image strip containing 60 lines from a tungsten halogen bulb filtered by a 674.8 nm central wavelength, 17.8 nm FWHM interference filter; below, the corresponding instantaneous spatial frequency.

Even within sixty lines, it can be seen that the fringes are not strictly vertical. This tilt represents a phase shift between successive line interferograms, and limits the maximum number of lines to be averaged to increase the SNR. The minimum number of lines is determined by the minimum SNR for which the IPSS algorithm is still able to recover the position of the phase step with reasonable false alarm rates. A more detailed study of the optimum number of lines, and of the mechanisms to deconvolve the optical distortion, allowing a larger number of lines to be averaged, was not performed within the scope of this research, and is indicated for future work.

The first target to be measured was the monochromator. Its diffraction grating was adjusted to a central wavelength of 674.8 nm, and the entrance/exit slits set to give a FWHM of 5.4 nm. The interference filter employed had the same central wavelength, and an FWHM of 17.8 nm. These settings aimed to make use of the target-filter combination with the highest detection sensitivity, in accordance with the studies done in chapter five. Sixty lines were averaged, between the lines 250 and 310 of the 576 available, in the central portion of the image. This number of lines was chosen by experimentation to be optimum for phase step detectability. An example of the application of the algorithm to a pair of images is seen in fig. 6.8. The image strip in the bottom of plot a) was acquired without the target. The one at the top had the target present. Plot b) presents the corresponding interferograms. The difficulty in seeing the phase change in the target-absent waveform may be an indication of averaging an excessive number of lines. Plot c) shows the instantaneous frequency signals. The target-absent signal (shown in red) displays almost no noticeable frequency changes. The target-present signal, however, shows two large frequency spikes, one of them possibly due to noise pick-up. Even so, a change in the period of the blue curve in plot b) can still be seen. In plot a) it is also interesting to notice that the two interferograms are in phase before the phase step happens (column no. 500), and in anti-phase after that column.

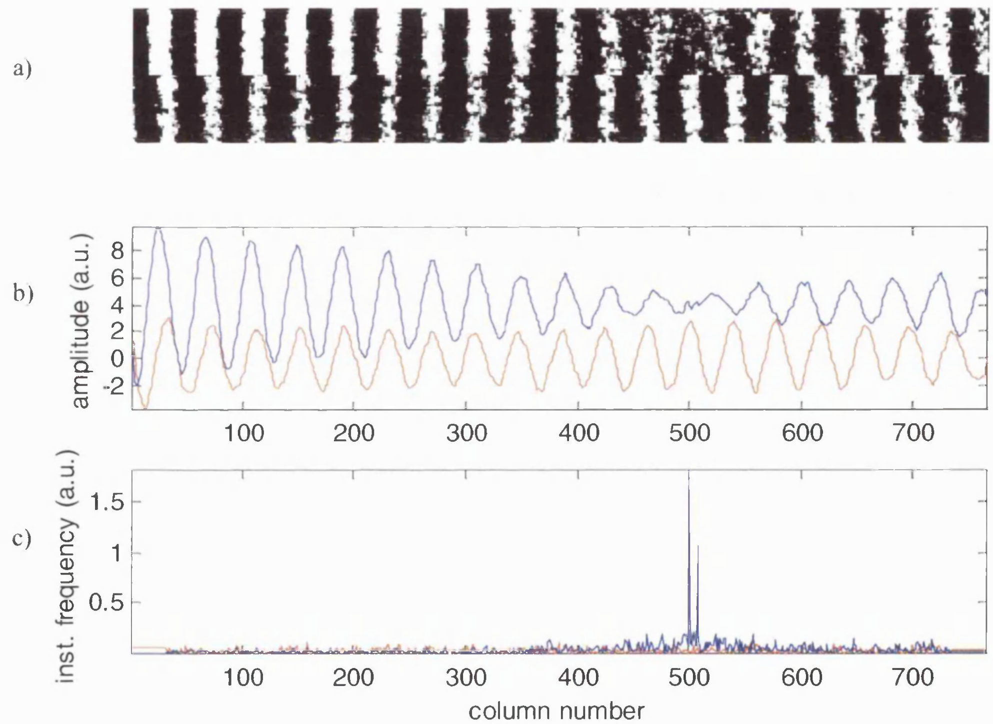


Figure 6.8 - Algorithm for recovery of the phase step from interferogram images. a): image strips. b): interferograms. c): instantaneous spatial frequency. In plot a) the lower image has no target, and upper image has a target. In plots b) and c), the red waveforms have no target, and the blue waveforms have a target.

The results of the application of the IPSS algorithm to the monochromator images are depicted in fig. 6.9. The conversion factor from pixels to microns was obtained by measuring the spatial period of one fringe and making it equal to one wavelength, 674.8 nm, only for comparison with earlier temporal plots. The MDSCR using the spatial approach, and the simple detection criterion presented in chapter five was –21.5 dB, corresponding to an input SCR of –33.4 dB.

The result in the system input is approximately 3 dB better than the one obtained using our temporal approach, shown in fig. 5.30 in chapter five. The average dispersion coefficient (standard deviation/average) of the measurements was 1.17, against 1.5 in the temporal case. Hence, with the monochromator as a target, the spatial approach is more sensitive and has less noise, at the expense of lower update rates, in this case limited by the frame rate of the camera (25 Hz in the interlaced mode). The fast piezoelectric transducer used in chapter five had an update rate of 500 Hz.

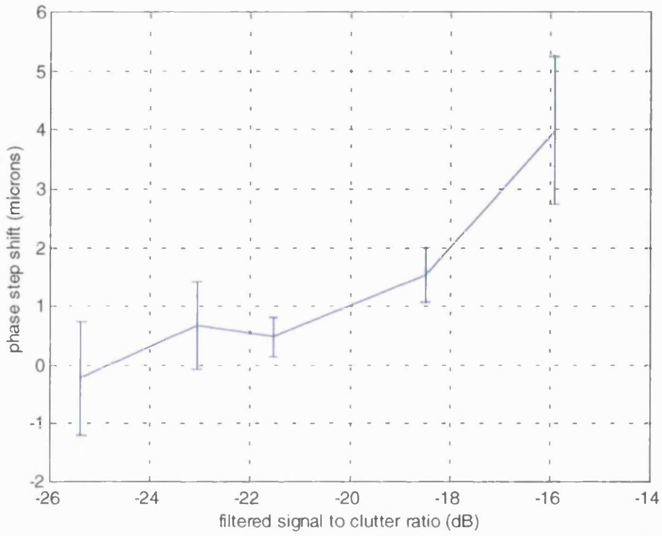


Figure 6.9 - Detection of a monochromator target from spatial interferograms

The next target used in the detection experiments was a 5 mW He-Ne laser, and the background was simulated by a 20 W tungsten halogen bulb, both used in chapter five. The interference filter was replaced by one with a FWHM of 11 nm, and a central wavelength of 632.8 nm. An objective lens with a magnification of 10 times was used to expand the beam (fig. 6.1). The signal to noise ratio was good enough to allow the recovery of useable interferograms averaging only 3 lines in the image. For the same reason, Wiener filtering was not required for noise removal. The objective lens in the laser path and the collimating lens in the tungsten halogen bulb path were adjusted to give coincident spots in the image plane, and the CCD array imaged a part of these two overlapping spots, indicated with a rectangle in fig 6.10. Although the target and background spot sizes were the same at the interferometer's entrance, the different degrees of collimation of these two light sources caused different spot sizes in the CCD array plane. In fig. 6.10 the laser spot is the brighter, and smaller than the background spot. The dimmer circular image to the left of the largest circle is a ghost reflection from the front surface of the interferometer's beamsplitter on the exit plane. The splitting surface of this beamsplitter is its back surface with respect to the interferometer's entrance plane (see fig. 6.2). Another interesting feature seen in this figure is the noticeable fringe phase shift at the upper right part of the image, between the background-only and target-plus-background regions of the image, demonstrating the feature (fringe phase shift in presence of a target) explored in IPSS.

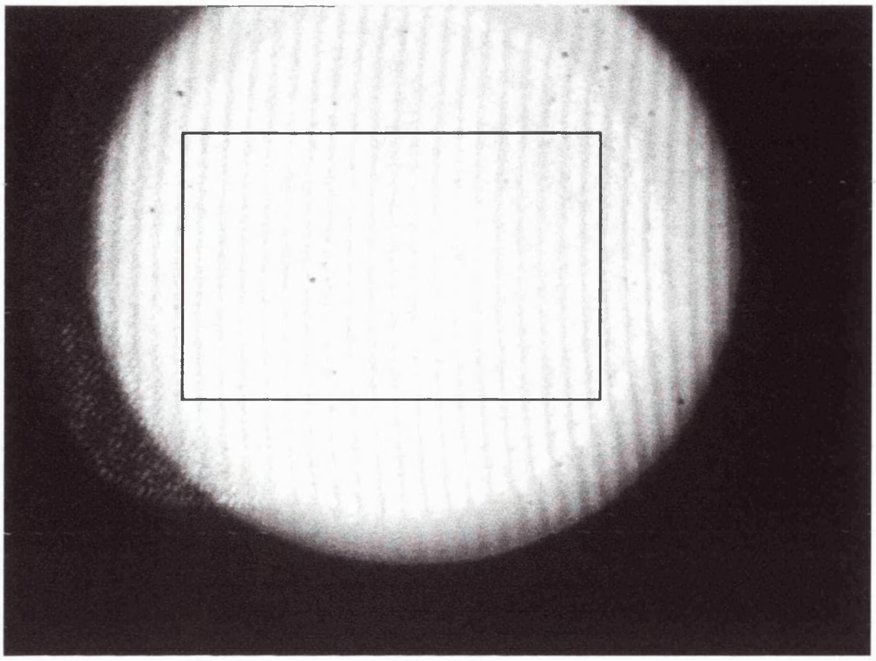


Figure 6.10 - Fringes from the filtered light of a tungsten halogen bulb (larger circle) and from the He-Ne laser (smaller circle), and region acquired by the CCD array (rectangle).

The algorithm used with the laser was similar to the one used with the monochromator, with the modifications described previously. The results of the application of the IPSS algorithm to the laser fringe images are shown in fig. 6.11. The curve deviates from the usual shape in its high-SCR end, where almost no additional shift is observed for an increase in SCR. It is believed that this effect occurred due to the need to use the edges of the image to record larger shifts, with more distortion than in the centre of the image. Other possible causes were the increase in laser speckle as laser power increased, and the difficulty of performing a correct removal of fixed pattern noise, using images recorded without the target, due to the long coherence length of the laser. The filtered MDSCR was –17.6 dB, corresponding to an input SCR of –36.6 dB. The latter number is slightly better than the one obtained with the monochromator, and is the worst MDSCR of the four techniques used in this thesis. The main reason for this relative performance degradation is speckle, a factor not found using the monochromator, nor in the experiments using temporal interferograms shown in chapter five. The amount of noise present is expressed by the dispersion coefficient, 6.65, the largest and worst of all detection experiments performed in this thesis. Although the performance with respect to SCR was not as good as in other experiments, the system's measured detection sensitivities are still high, adding to the advantage of using no moving parts.

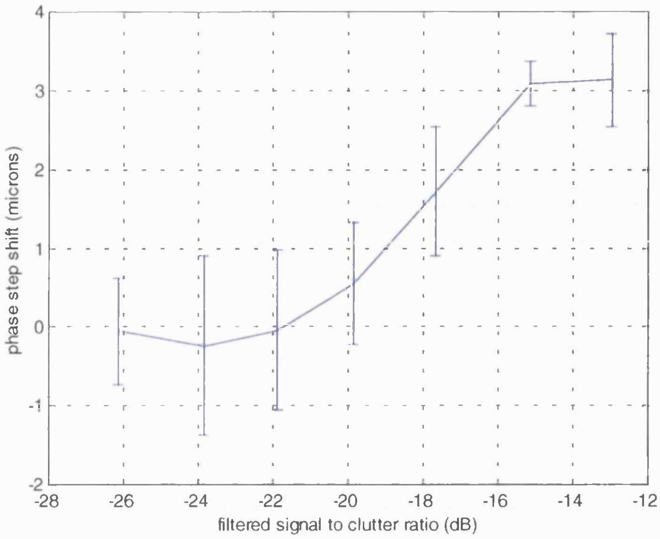


Figure 6.11 - Detection of a laser target from spatial interferograms

6.3. A coherence-based imaging spectrometer

This section presents the application of the IPSS algorithm to form an image whose pixels' greyscale intensity is proportional to the degree of coherence of the light radiated by the point in the scene imaged by that pixel. A brief review of previous work on this subject is included, and experiments using these other approaches are replicated for comparison purposes.

In his doctoral thesis, Sutton [Sut82] employed a method to detect a laser emission in a daylight background that used a Fresnel biprism to generate interference fringes, and a scanning mirror to create modulation. The method was reviewed in chapter two. A glass plate is added to one of the sides of the biprism to create a bias path difference, making a high pass temporal coherence filter, as only light with a coherence length longer than this bias would cause interference. The detection is based on the presence of a modulated signal, which is recognised by an electronic circuit comprising a tuned amplifier, acting as a bandpass filter, an envelope detector and a threshold comparator. No SCR figures are given, but the sensitivity of the approach is demonstrated by the spatial location of the emission from a 0.5 mW He-Ne laser at a distance of 560 m, in a one-dimensional signal. Although the low source power and the target-sensor distance indicate that high sensitivity was demonstrated, the thesis did not include a two-dimensional (or imaging) demonstration. Hickman [Hic88] [Hic89] has used a frame subtraction method to recognise the modulation generated by a laser in the same situation studied by Sutton, consisting of frame subtraction. As described in chapter two, this work, called by the author "an initial demonstration of imaging" [Hic89] used only two frames and claimed the detection of a laser in an image with an SCR of –20 dB.

In this thesis, experiments were carried out to replicate the work mentioned above under controlled, laboratory conditions. This was thought to be better than relying on published information and making assumptions about parameters not informed in the publications. The experiments consisted of the measurement of the smallest signal-to-clutter ratio (SCR) for which the target could be located (for a point target) or recognised (for an extended target) on a screen. The target was a 633 nm, 5 mW unpolarised He-Ne laser and the clutter was a light source comprising a 20 W tungsten halogen bulb with a blackbody temperature of 3220 K, and collimating optics (both used previously). The light from the two sources were combined in a cube beamsplitter, and sent to the detection system, as shown in fig. 6.1. The target radiation was spatially distributed in two different ways: a point target, represented by the laser beam without expansion, and an extended target, formed by expanding the laser beam, and passing it through a USAF 1951 test chart. Variable SCRs were obtained by inserting a variable attenuator and neutral density filters in the laser path to reduce its transmitted power. Due to this discrete SCR variation, only a limited number of SCRs were possible, varying from six to ten.

The first experimental attempt consisted simply of recording an image from the combined target/background radiation and trying to find an object in it. Fig. 6.12 displays a set of images acquired using a frame grabber (Matrox Pulsar) and a monochrome CCD camera (Sony XC77RR-CE) coupled to a zoom lens (Computar 501 4377) pointed towards the exit surface of the cube beamsplitter used to combine the light from the two light sources. No interferometer was used. The numbers in dB are the SCRs measured as follows: a photodetector was placed behind the iris, replacing the interferometer seen in fig. 6.1. The powers from target and background were measured separately, and a correction factor of +4.78 dB was applied, to account for the fact that the target image does not fill every pixel in the image, as it passes through the test target mask. This factor corresponds to the total number of pixels in the image (576 x 768), divided by the number of target pixels, to find the "per pixel" SCR. It allows the comparison of results from imaging and non-imaging experiments, which represent a single pixel detection system. Fig. 6.12 shows the images acquired from the target and background described, with different SCRs. The image in the bottom right of the figure is the test chart. Using the ability to see spatial structure in the images as a detection criterion, the image corresponding to SCR=+1.7 dB has the lowest SCR detected, and this SCR, from now on, is defined as the sensitivity of this visual detection system. The images did not receive any processing.

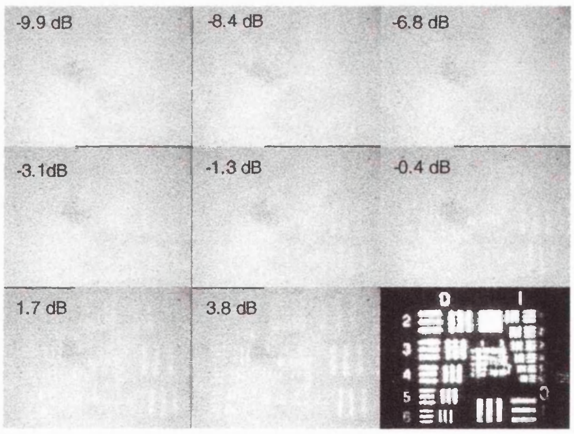


Figure 6.12 - Images combining target and background with different SCRs

Following this initial experiment, a narrowband interference filter centred on the 633 nm He-Ne line and with a FWHM of 11 nm was placed in front of the CCD camera. A similar set of images was acquired, and the measured sensitivity (again using the eye) was now +2.6 dB after filtering, corresponding to −16.6 dB at the input of the detection system (the filter introduced a gain of 19.2 dB, calculated by the ratio of white light powers measured with and without the filter). Hence, the inclusion of spectral filtering increased the sensitivity by 18.3 dB.

Removing the spectral filter, the images shown in fig. 6.12 were used in a simple algorithm, consisting of thresholding the images with a threshold selected by observation of the image amplitude histogram, and plotting a binary image with the white pixels corresponding to threshold crossings. Fig. 6.13 shows the same data used in fig. 6.12, thresholded at 84%. The spatial structure from the test chart bars can be seen only at the SCR of 3.8 dB, worse than in the unprocessed data. Inserting the narrowband filter, the results are the same as in the unprocessed case (+2.6 dB). Hence, the use of a threshold reduced the sensitivity when not using spectral filtering, and did not change it when using spectral filtering.

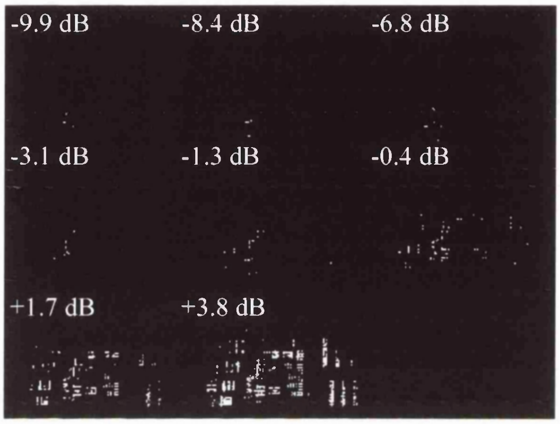


Figure 6.13 - Image with the same light input as in fig. 6.12, thresholded at 84%

The next experiments used the whole arrangement of fig. 6.1. One of the interferometer's mirrors was moved longitudinally to create a path difference between the two beams. The movement (of the translation stage) is controlled by a motor controller (Klinger MC4), according to the commands generated in a computer running a program in Labview language. The radiation exiting the interferometer is collected by the zoom lens and coupled to the CCD array, whose video signal is digitised in the frame grabber and stored for off-line processing in the computer. The frame grabbing process is synchronised to the mirror motion by triggering the frame grabber from the motion controller as described in section 6.1. As the path difference is changed through mirror displacement (not mirror tilt as in the previous section), interference fringes are formed in the interferometer exit plane, provided that the coherence length of the light source is longer than the path imbalance [Hec98]. The path difference is biased to a value much longer than the coherence length of the background, the path difference is varied and successive images of the target-background combined scene are recorded. Due to its longer coherence length, the radiation from the He-Ne laser will produce fringes [Sut82], which will be measured on the CCD sensor as a temporal modulation. An observer at the interferometer exit plane will see the target blinking, and no change in the background clutter, apart from noise. Scan time must be long enough to allow the recording of fringes enough for the modulation to be recognised. Frame rate must be higher than twice the modulation frequency (Nyquist criterion [Ope89]).

One simple way of detecting the coherent target is to calculate the pixel-by-pixel difference between successive frames, and look for time-varying intensities. It was attempted to do this initially with two image frames of a three bar target at an SCR of –11.6 dB, similarly to the work reported in [Hic88], but the target was not detected due to the increase in high frequency noise, caused by the differentiation operation. This SCR was chosen because, using other techniques, the same target has been detected. In order to compare this experiment with the one reported in [Hic88] and [Hic89], some other parameters must be considered. In both cases, the two frames were acquired asynchronously, after a time interval much longer than the standard TV frame rate (25 Hz). Concerning path difference imbalance, Hickman has used 100 μm, and we have used 30 μm. Considering a background coherence length of 1 μm (measured in chapter four) and a laser coherence length of 20 cm (given by the manufacturer), and using eq. 2.4, which we reproduce here:

$$R = \frac{\delta I_S}{\delta I_b} = \exp\left[\frac{\pi^2 L^2}{2}\left(\frac{1}{L_{cb}} - \frac{1}{L_{cs}}\right)^2\right]$$

In this case, the source coherence length L_{CS} is much longer than the background coherence length L_{CB}, and the term in brackets is dominated by $1/L_{CB}$. However, as the path difference imbalance is also much longer than LCB in both Hickman's work (100 μm) and ours (30 μm), the theoretical rejection ratio R [Sut85] would be very high in either case. The sensitivity then is limited by background residual modulation, affecting both cases equally. Therefore, in similar experimental conditions, the result claimed by Hickman [Hic88] (detection at –20 dB in two frames) was not reproduced.

In [Hic89] a possibly more robust method was used, consisting of taking a longer frame sequence sampled at 25 Hz. Images of a target with a FWHM of 14 nm detected in a 200 nm spectrally filtered background are included, but no sensitivity figures are given. In order to reproduce such experiment (although with different target and background) here, a data cube with two spatial and one temporal dimension was created. A computer program in Matlab language was written, to differentiate the time signal for each pixel, find the maximum value of the result attempting to find the modulation from the target and convert the resulting matrix to a greyscale for visualisation. Fig. 6.14 presents one of the frames of the 20 frame long original sequence, where the target cannot be distinguished, and the processed image, where the target, a set of three vertical bars found in the bottom-right corner of the USAF test chart shown in fig. 6.12, appear brighter than the background. The MDSCR in this case was –11.6 dB, the same unsuccessfully used in the two-frame experiment, but better than the figures from the thresholding experiments.

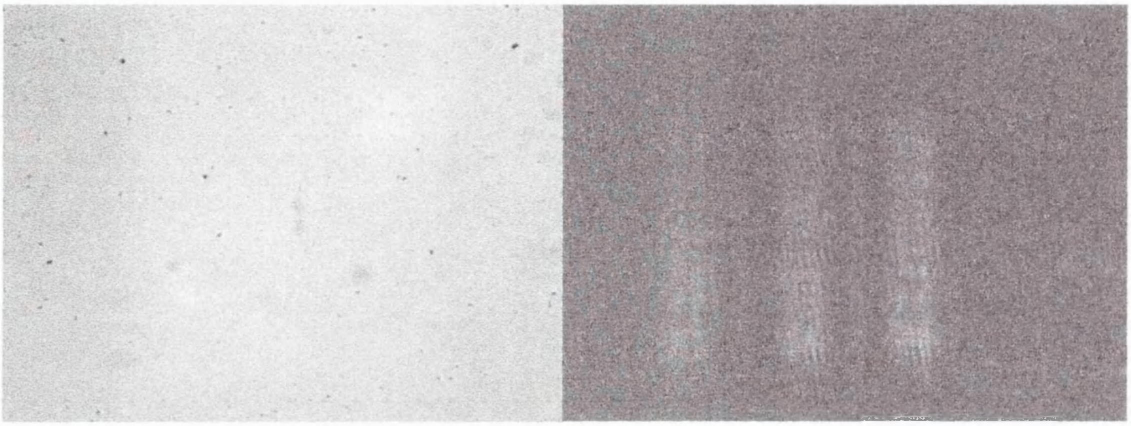


Figure 6.14 - Frame differentiation method, showing the raw image to the left, and the processed image to the right.

We have improved the MDSCR of the maximum amplitude method just described to −17.0 dB by calculating the standard deviation, instead of the maximum, of the differences between successive points in the time signal. This was found necessary because the use of the maximum was too susceptible to noise pick-ups in background pixels of the images. By applying the standard deviation method to the detection of an unexpanded He-Ne laser beam in the same background, we have obtained a MDSCR of −4.88 dB. This number was obtained after the application of a correction factor of 33.2 dB, used to calculate the per-pixel SCR, and is worse than the one obtained with extended targets. This issue is discussed in section 6.4. Appendix III presents a movie of the variation in intensity of the three bars as the path difference is changed.

The frame-differentiation method previously described is limited by the amount of noise generated by the differentiation process. The noise variances of successive frames tend to add, and intensity variations actually due to noise are detected by the difference algorithm, preventing better sensitivities from being obtained. The method used by Sutton [Sut82], consisting of narrowband filtering at the frequency of the expected modulation, does not present the drawbacks of frame differentiation, and was reproduced here in a set of images of an unexpanded He-Ne laser (point target) in the same tungsten halogen bulb background described previously. The software algorithm, written in Matlab language, initially subtracts a reference image with no target from all frames in the sequence, forms a three dimensional data cube (as previously described), reorganizes the data in one time sequence per pixel, and bandpass filters each time sequence, using a carrier frequency calculated from the wavelength of the radiation and the resolution of the translation stage (0.1 μm per step). An example of the resulting sequence is shown in fig. 6.15. The coherent target clearly produced higher modulation indices than the background. The amplitude of the signals is reduced at their initial samples due to the lack of a proper initial condition to the difference equation that implements the software filter, difficult to calculate from the short length of the sequence. It can be seen that the residual background modulation is in anti-phase with the target, the reason of this behaviour being unknown. An image is then generated from the greyscale intensities, made proportional to the amplitude of the modulation. By using this approach, a minimum detectable SCR of –4.88 dB was obtained, as in the case of frame differentiation, but with the advantage of detecting a point target. Fig. 6.16 a) shows an image obtained using the modulation recognition method [Sut82], whose size had to be reduced due to computer memory limitations. The target can be seen as the white spot at the centre of the image. Appendix III presents a movie of the laser target blinking, while the background stays constant, apart from noise. In fig. 6.16 b) the modulation depth in a horizontal strip of one line passing through the target is shown, locating its position in the horizontal dimension, the same way as Sutton has done in his doctoral thesis [Sut82].

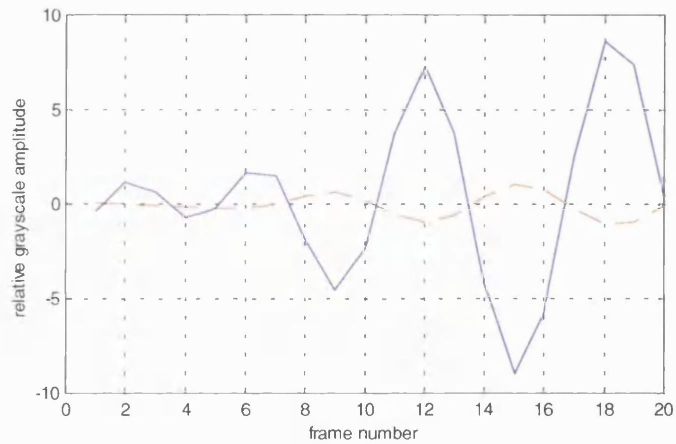


Figure 6.15 - Modulation caused by a coherent target (blue) and the incoherent background (red)

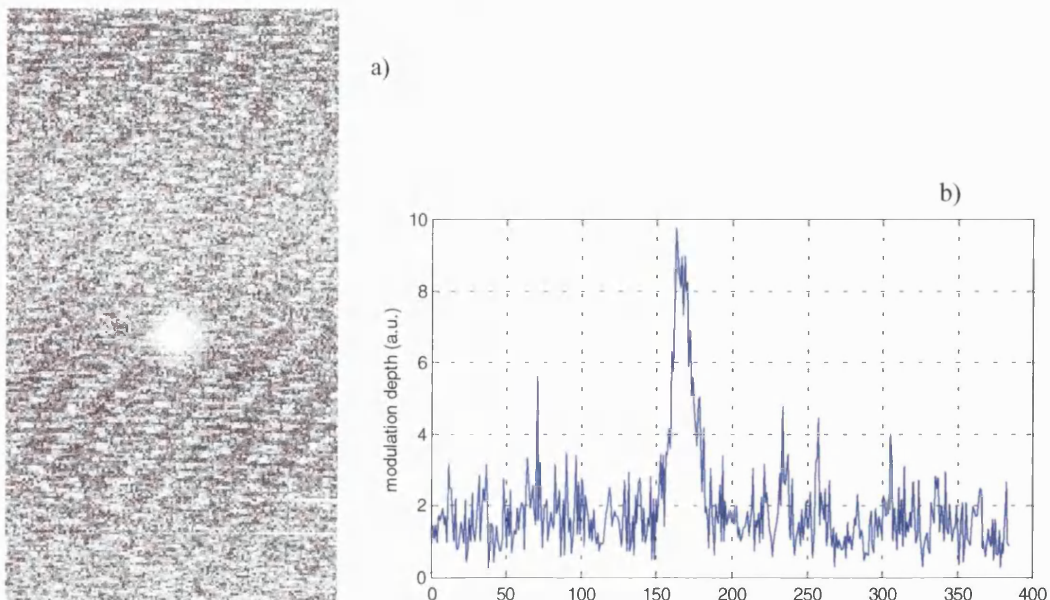


Figure 6.16 - a) - A laser spot (in the centre of the image) detected by modulation recognition; b) - Modulation depth in a horizontal line of the image in fig. 6.15 a)

Because of the fringe format shown later in fig. 6.17, which was due to distortion in the optical components, it was not possible to enlarge the fringes, so that a single circular fringe would be seen in the field of view of the CCD camera. Hence, pixels in different regions of the image did not blink in phase. Therefore, the modulation recognition method did not work for extended targets, as the different phases and even frequencies of the modulation throughout the image caused different output amplitudes in the bandpass filter used by the method.

After replicating the experiments performed by previous workers, the IPSS algorithm was applied to the detection and location of the same laser target in a two-dimensional image. In the temporal mode of the technique, rather than tilting one of the interferometer's mirrors, it is scanned longitudinally as happens in FTS, but limiting the length of the scan to the region of the region of interest in the interferogram, where the phase step happens. In this mode, the interferometric system described in section 6.1 employed the components shown in fig. 6.1, with the addition of an interference filter after the iris, with a mean wavelength of 632.8 nm (coincident with the He-Ne laser line), and a FWHM of 11 nm. The target, either an expanded beam through the USAF test chart or an unexpanded beam, was dimmed by a combination of neutral density filters and a variable attenuator. By changing this combination, three different SCRs for the point target, and four SCRs for the extended target were obtained, and a sequence of 500 frames were acquired for each SCR, for both extended and point targets. In order to record the movie, the automated image acquisition system was programmed to grab one frame, store it in the computer's memory, move the translation stage 0.1 μm, wait for the stage's position to stabilise, and grab another frame, at a rate of 7.5 frames-per-second. The path-difference sampling rate was 0.2 μm per step (twice the stage displacement). This sampling rate, the smallest possible with the available stages, was enough to obey the Nyquist sampling criterion, with a wavelength of 0.633 μm, although some degradation in the detection of the phase transition did happen due to this sampling limitation, as the phase step possesses higher frequency content than the fringe carrier (shown in chapter seven). The tilting mirror of the interferometer was adjusted to give fringes as close to circular as possible, as the approach requires an equalisation of phases throughout the image, but only hyperbolic fringes (shown in fig. 6.17, with the hyperbola's focal points out of the image) were obtained due to the imperfections in the mirror's mounting mentioned in chapter five. The algorithm in this case consisted of the following steps:

a) loading 500 frames with 288 x 384 pixels each;
b) reducing the original data volume (by discarding samples) from 288 x 384 x 500 to 140 x 114 x 180, in order to allow processing by the computer available;

c) reorganising the samples to transform the 180 images obtained in b) into a data cube containing 140 x 114 interferograms (two spatial dimensions), with 180 samples each (temporal dimension);

d) subtracting a reference image, pixel-by-pixel, to remove the background and fixed pattern noise, as done in section 6.2;

e) applying the IPSS algorithm (described in chapter seven) to each of the 140 x 114 temporal signals formed, and finding one path difference value for the phase step, per signal;

f) subtracting, from the values obtained for each pixel in e), the value obtained by applying the same algorithm to a region of the image without the target, which is the no-target reference;

g) creating a greyscale representation and a binary image of the shifts in the phase step. The greyscale is calculated linearly from 1 to 258 (8-bit), level 1 for the darkest pixel and level 256 for the brighter. This algorithm is pictorially depicted in fig. 6.18. Its function named "average mask" is explained later in this section.

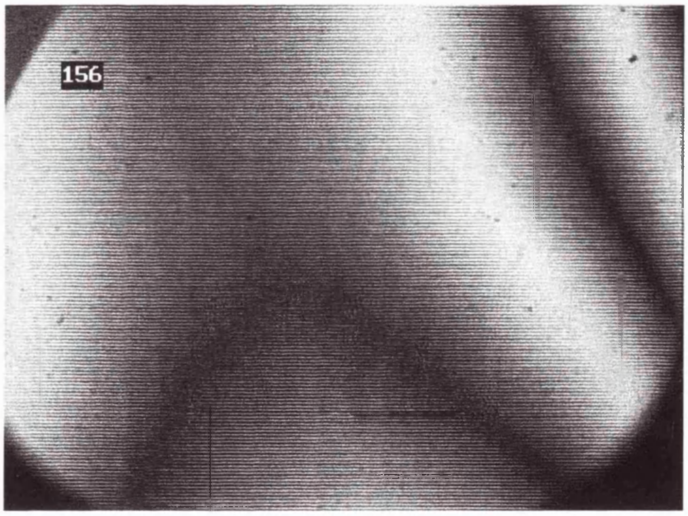


Figure 6.17 - Fringe format in the IPSS imaging experiments. The fringes were formed in the bottom-centre and moved to the top-left corner of the image (a movie is included in appendix III)

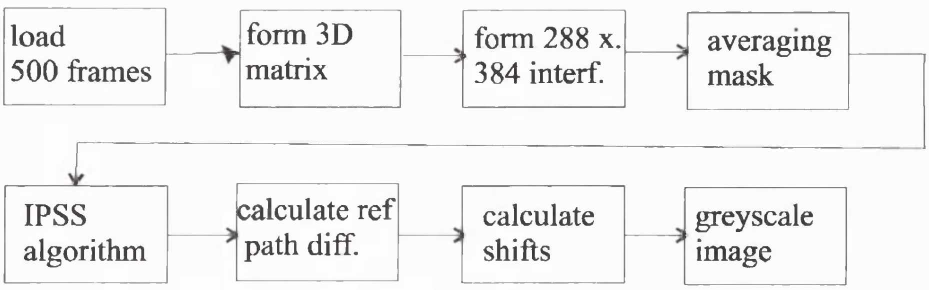


Figure 6.18 - Imaging algorithm block diagram

The algorithm described above was run with the sequences recorded with a point target, and no detection occurred, due to the signal-to-noise ratio being too low. The region of the interferogram to be investigated is the one with the lowest visibility, creating a weak interferogram signal to compete with noise. The solution was to apply a 3 x 3 averaging window, which calculated the average of the interferograms (sample by sample) from nine neighbouring pixels, and replaced the central pixel with the result. The new size of the image was then 128 x 96 pixels. This averaging operation, a spatial filtering technique used in image processing [Wee98], reduced the temporal noise measured by each pixel and allowed the detection of the point laser target at an SCR of -22.26 dB at the input of the detection system, the best figure obtained in this chapter for a point target. Fig. 6.19 shows the interferogram generated from one pixel in the image before averaging and fig. 6.20 presents the interferogram obtained after the averaging operation, now with the familiar shape seen previously in this thesis.

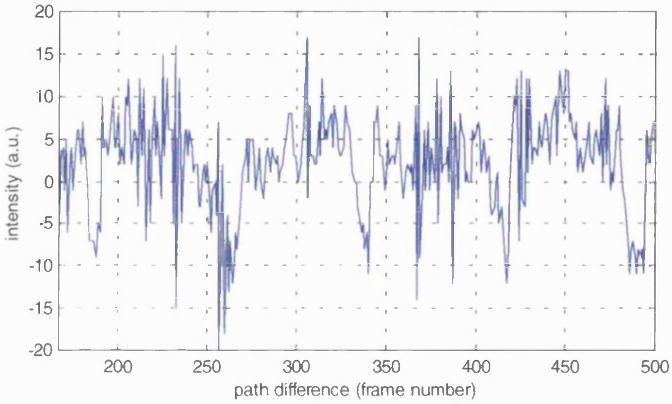


Figure 6.19 - Interferogram from frames 167-500 without averaging

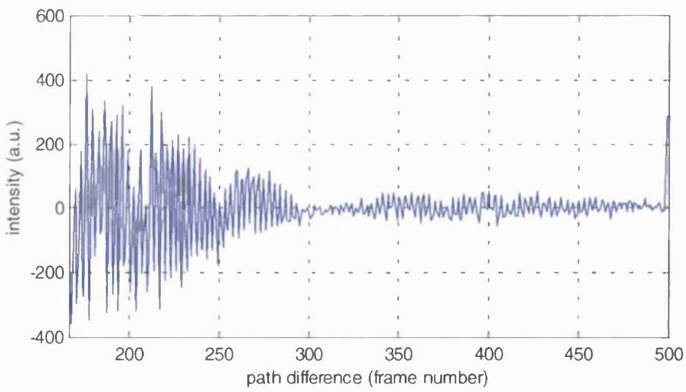


Figure 6.20 - Interferogram from frames 167-500 averaging in a 3x3 window

Fig. 6.21 shows one of the frames of the original sequence, showing the shape of the fringes, the greyscale image calculated by the algorithm, and the binary image obtained with a threshold of 0.99. The greyscale image is pixelated due to the reduction in spatial resolution, caused by the averaging process. In the binary image, the light coming from the laser was detected as a white spot, in the correct position where the emission was initially located. If a lower threshold had been used, detection would still have been possible, although with false alarms, as shown in fig. 6.22. In this figure, calculated with a threshold of 0.85, a cluster of detections is seen in the region where the target was placed, and one false alarm is seen to the left of the cluster.

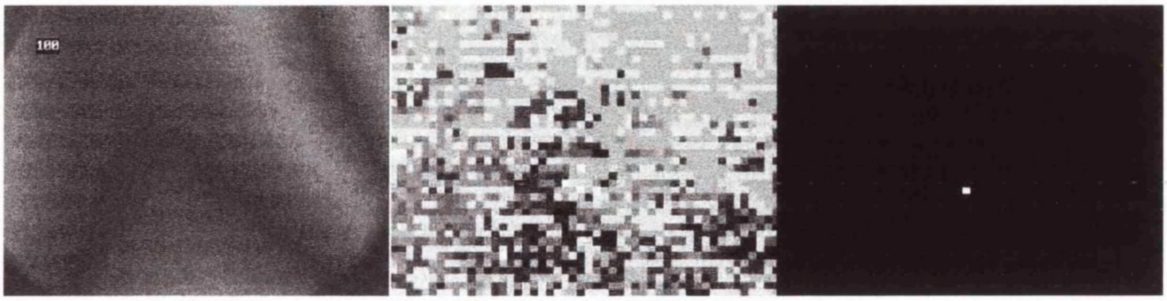


Figure 6.21 - Application of the IPSS algorithm showing, from left to right, raw, greyscale and binary images (threshold of 0.99).

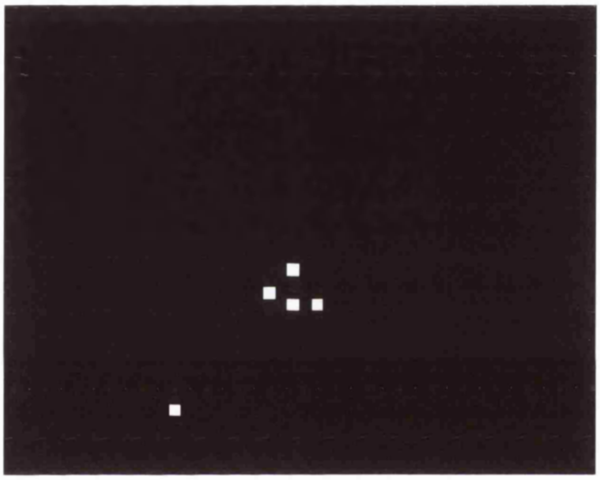


Figure 6.22 - Binary image of the result of the IPSS algorithm, with a threshold of 0.85.

As happened with the modulation recognition method, the detection of extended targets using the IPSS method was not possible, due to the phase delay that exists between different regions of the image, as the fringes travel from the bottom-right to the top-left (see movies in appendix III). This phase delay prevents a single reference path difference (the position of the phase step in the path difference axis without a target) from being obtained.

6.4. Conclusions

In this chapter, two sets of experiments combining the IPSS with the use of a CCD camera were presented. The first set had the interferometer employed to generate spatial interferograms, that is, interferograms in which the path difference was represented by pixels in horizontal lines of the imaging array. Fig. 6.23 presents a comparison of the results obtained in section 6.2, the detection of a He-Ne laser using the spatial technique, with the results obtained in chapter five of the detection of the same target. The results are in very good agreement, demonstrating that the physical phenomenon has been correctly measured, as the measurements taken using two completely different techniques coincide within their error bars. It was not possible to measure the behaviour with the CCD camera for SCRs higher than –13 dB, because of the distortion, laser speckle and difficulties in removing fixed pattern noise, as mentioned previously. The MDSCR of –36.6 dB (laser, spatial technique) is approximately 10 dB worse than the number obtained with the temporal approach, thus sensitivity is being traded off against the advantage of having no moving parts. While the temporal approach is suitable where very high sensitivities are required, the spatial approach is recommended in applications where reliability and robustness are essential.

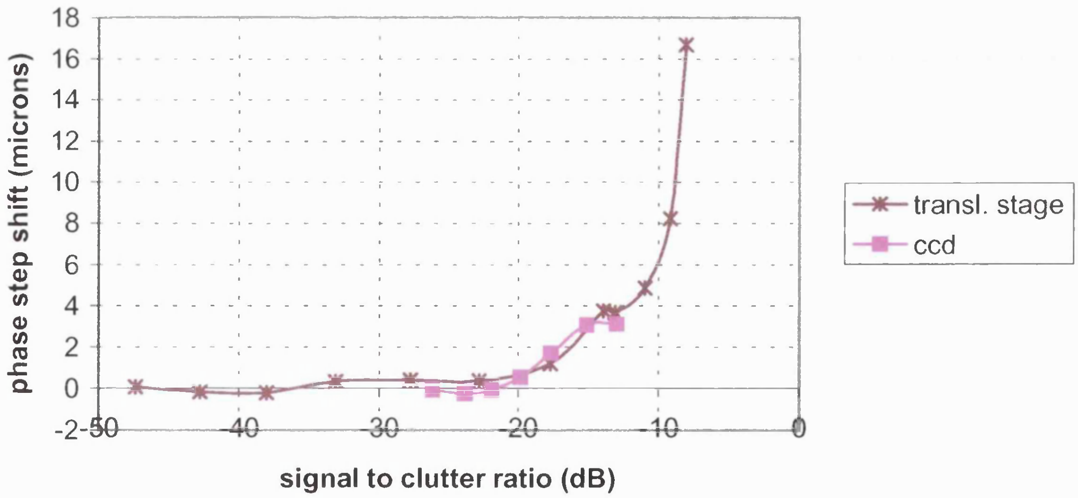


Figure 6.23 - Comparison of the results obtained using the CCD camera with the results obtained with a translation stage, shown in fig. 5.8

A partially coherent source formed by a grating monochromator was also used as a target in this chapter, and was detected (using the simplified error bar criterion) at an input SCR of –33.4 dB, ≈ 3 dB better than the sensitivity measured using the fast piezoelectric transducer studied in chapter five. Fig. 6.24 shows a comparison between the results obtained when detecting light from the monochromator, using the CCD camera, and using the translation stage (same data plotted in fig. 5.14 for BWR = 0.303). The agreement is only modest in between – 22 and -15 dB, but still reasonable if one considers the error bars on both curves, not displayed for clarity. As happened with the laser, higher SCRs could not be measured for the same reasons (except laser speckle). In both cases, the monochromator had the same central wavelength (674.8 nm) and FWHM (5.4 nm), and the same interference filter was used (same central wavelength, FWHM=17.8 nm).

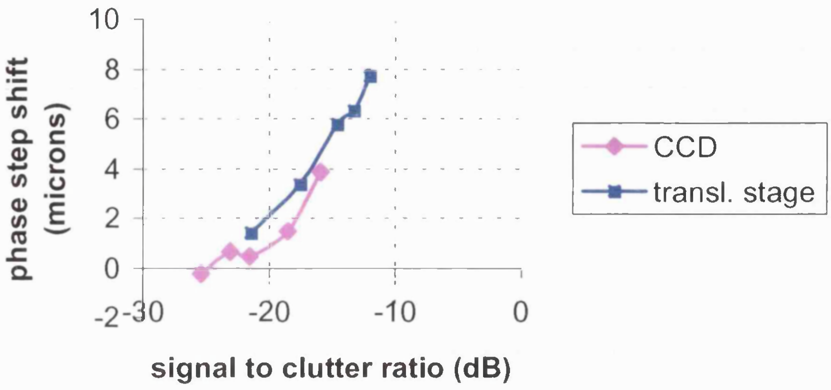


Figure 6.24 - Comparison of results obtained for a monochromator, using the CCD camera, and the translation stage. The displayed SCRs are given after filtering.

Another set of experiments comprised the extension of the IPSS approach to imaging using the coherence contrast between target and background. Table 6.1 summarises the results obtained. In general, the MDSCRs obtained by using point targets are worse than those obtained by using extended targets (after the correction used to obtain the "per pixel" SCR). When using human vision as the detection criterion, the higher detectability of extended targets than of point targets is explained by the larger number of points an extended target gives the eye-brain set of an human to integrate and make a decision. Comparing the different techniques employed, for point targets, the most sensitive approach is IPSS, with an advantage of 17.4 dB compared to modulation recognition and frame differentiation, which performed similarly. When using extended targets, IPSS and narrowband filtering did not work, and one can only say that frame differentiation is slightly better (lower MDSCR) than optical filtering, with the advantage of wideband operation. The problems found with imaging extended targets using IPSS were due to the poor quality of the optical components employed in the experiments and alignment difficulties.

To the best of the author's knowledge, the work presented in this chapter represents the first comprehensive demonstration of imaging based on the coherence properties of the target, in contrast to the (lack of) coherence of the background. Low coherence interferometry, and its associated techniques such as optical coherence tomography (see chapter two) explore the coherence properties of the light source, not of the target. However, the studies shown in this chapter comprised only stationary targets. The application of the IPSS technique to the imaging of moving targets is out of the scope of this thesis, and is an indicated direction for future work.

Experiment	Minimum detectable SCR (dB)	Number of frames required	target
No processing, without optical filtering	+1.7	1	Laser, extended
No processing, with optical filtering	-16.6	1	Laser, extended
Thresholding without optical filtering	+3.8	1	Laser, extended
Thresholding with optical filtering	-16.6	1	Laser, extended
Frame differentiation	-17	20	Laser, extended
	- 4.9	20	Laser, point
Modulation recognition	-4.9	20	Laser, point
IPSS - temporal mode	-22.3	180	Laser, point
IPSS - spatial mode	-33.4	1	Monochromator, non-imaging
	-36.6	1	Laser, non-imaging

Table 6.1 - Imaging experiments results summary

Chapter 7

Algorithm Development and Comparison

The Interferogram Phase Step Shift (IPSS) technique has been described as a whole in chapter three. This chapter describes the IPSS algorithm, a novel part of that technique which we have invented. The IPSS algorithm's main function is to perform phase demodulation of the interferogram signal to locate a frequency spike in it, whose displacement is used to detect the presence of a target. The main purpose of this chapter is to justify the choice and setting of parameters of the algorithm employed for detection, which operates in the interferograms recorded using the methods described in chapters five and six. The results of this analysis served as a tool for the design of the algorithm. The chapter starts with an analysis of the experimental data generated during the research, which comprised temporal interferograms, images of line fringes used to form interferograms in the spatial domain, and sequences of images used to create movies. The candidate techniques are then chosen, described and compared by constructing their Receiver Operating Characteristics (ROC), and it is shown that the IPSS algorithm has an advantageous performance over the other techniques studied. A conclusion session finishes the chapter characterising the performance achieved by the IPSS technique in terms of minimum detectable signal to clutter ratio (MDSCR), probability of detection, and probability of false alarm, a more rigorous characterisation than the one done in the previous chapters (MDSCR only). These results are compared with detection benchmarks from the literature. The computer programs described in the chapter are listed in appendix I.

7.1. Analysis of the experimental data

It is the purpose of the IPSS technique to detect the presence of an object, radiation or absorption in the field of view of the sensor via the analysis of a small section of an interferogram with the characteristics described in chapter three. These interferogram sections are signals with amplitude and phase (hence, frequency) varying with time. An algorithm using the IPSS technique to perform the detection task must then measure, as accurately as possible, the incremental change in the coherence of the incoming radiation, by measuring the displacement of a null in its complex degree of coherence.

However, many different effects corrupted the real signals acquired using the techniques described in the previous two chapters. Noise from the electronics and detector prevented the existence of a null in the interferogram, besides creating false phase changes. Imperfections in the optics reduced the amplitude of the fringes due to the inability to form circular fringes, and the micropositioning equipment created an artificial displacement of the interferogram, due to its lack of repeatability. In some cases, lack of linearity in the movement of the interferometer's mirror generated a chirp effect in the interferogram that was a change in fringe frequency, an experimental artefact capable of jeopardising any detection technique relying on that frequency. The chirping effect is shown in fig. 7.1. Plot (a) depicts an example interferogram signal measured using a piezoelectric transducer, where the period of the fringe carrier is seen to vary considerably across the waveform, due to transducer non-linearity. In plot (b), a spectrogram shows the variation of frequency (vertical) with time (horizontal). The frequency increases from top to bottom, being relatively low at the start of the waveform, higher at its end, passing through a high frequency transition in the region of the phase step (~ 0.07 s), shown as the yellow strips extending up to the top of the plot. This frequency increase is the phenomenon explored by IPSS. A suitable algorithm, then, must cope with those effects and produce minimum dispersion (defined here as standard deviation/average) when measuring the null displacement. As mentioned previously, in order to reduce the severity of these artefacts, the experimental signals acquired either with the translation stage, piezoelectric transducers or CCD camera were recorded in pairs, and five pairs were acquired for each target strength (SCR).

Appendix II presents a compilation of the experimental data acquired during the research, containing details of the light sources used as targets and backgrounds, and settings of the experimental arrangement, such as sampling frequency, analogue gain and cut-off frequencies. In order to facilitate the analysis of the data available, around 1500 interferograms were arranged in 44 categories, and 3600 image frames were separated in seven categories..The large amount of data includes alignment, calibration and effective runs, as well as the actual measurements of interest. As the analysis of the whole data set would be too time-consuming, twenty-five groups (twenty-one of temporal interferograms and four of images), the most representative of the various experiments performed, were selected from the initial set to be analysed and used as input data to the tasks of algorithm development and optimisation. Some of these groups of data were used to derive algorithm constants, while some others were used to test them.

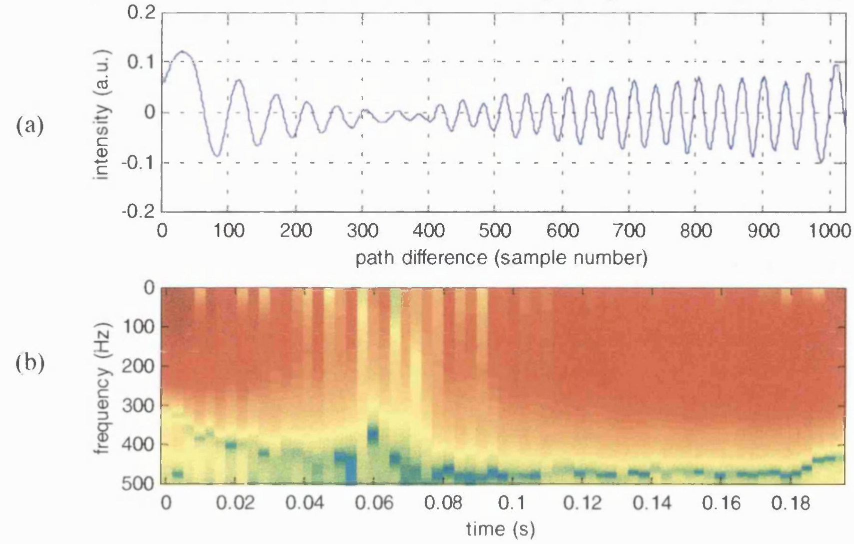


Figure 7.1 - A spectrogram showing the chirping effect due to the non-linearity of the piezoelectric transducer. (a) - Interferogram signal. (b) - Spectrogram showing the variation of frequency (vertical) with time (horizontal).

In order to aid the design of software filters to be used in the detection algorithms described in the next section, a spectrum was calculated for one sample from each of the groups, using a Fast Fourier Transform (FFT) with a Hamming window [Ope89] to remove the undesirable effects of reversal in the direction of movement of the mirror, close to the edges of the interferograms. The interferograms are multiplied sample by sample by the window, which emphasises the central portion of the waveform and reduces the amplitude of its edges. The FFT had 2048 points, and the complex coefficients were multiplied by their conjugates to give the power spectral density (PSD). Fig. 7.2 presents the power spectral density corresponding to the first group, calculated from data acquired with the slow piezoelectric transducer. In this spectrum, it is easy to recognise the fringe signal at 160 Hz, with supply noise at 100 Hz and some DC combined with the very low frequencies arising from the long-term amplitude variations in the interferogram. The 100 Hz appears because the light bulb flickers with the modulus of the supply, hence two times per cycle. The DC level is the residue of the manual adjustment of the ground level in the oscilloscope, and must be removed by software to increase the dynamic range. This spectral structure is similar for all waveforms generated by different methods, with variations in fringe frequency and in the presence of noise peaks.

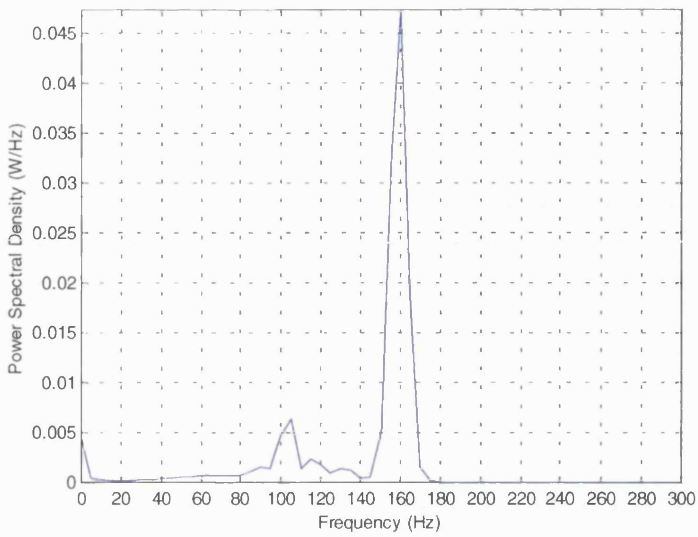


Figure 7.2 - Power spectrum of an interferogram acquired with the slow piezoelectric transducer

In the waveforms acquired with the fast piezoelectric transducer, of which a sample spectrum is shown in fig. 7.3, the fringe frequency is 45 kHz, far away from the 100 Hz noise peak, and no software filter is actually required. Nevertheless, the use of such a filter is mandatory when using the translation stage mirror-driving scheme. This is shown in fig. 7.4, which presents a spectrum calculated from data acquired using this method. Several noise peaks occur at frequencies up to 350 Hz, most of them stronger than the signal of interest, calculated to be at 470 Hz, using eq. 4.4. By switching off the stepper motor of the translation stage it was found that the noise peaks were caused by the motor and represented the dominant noise source in this case. The 100 Hz noise is almost unnoticeable in fig. 7.4. The stronger presence of noise in the interferograms acquired with the translation stage was observed in chapter five through the calculation of the dispersion coefficient of the measurements.

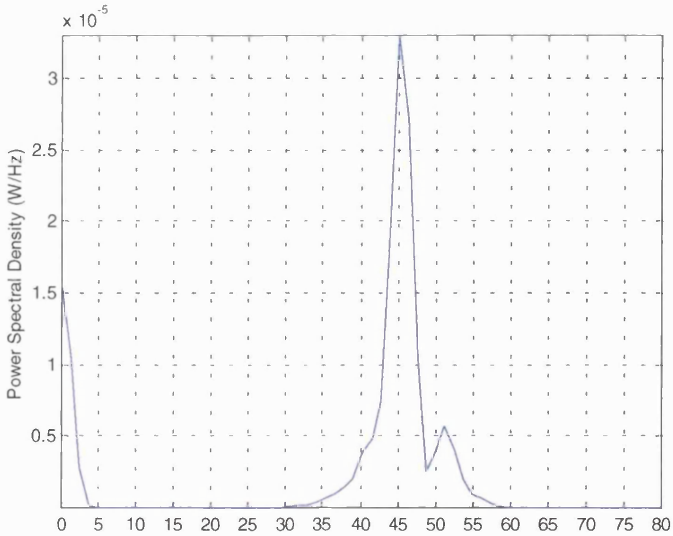


Figure 7.3 - Power spectrum of an interferogram acquired with the fast piezoelectric transducer

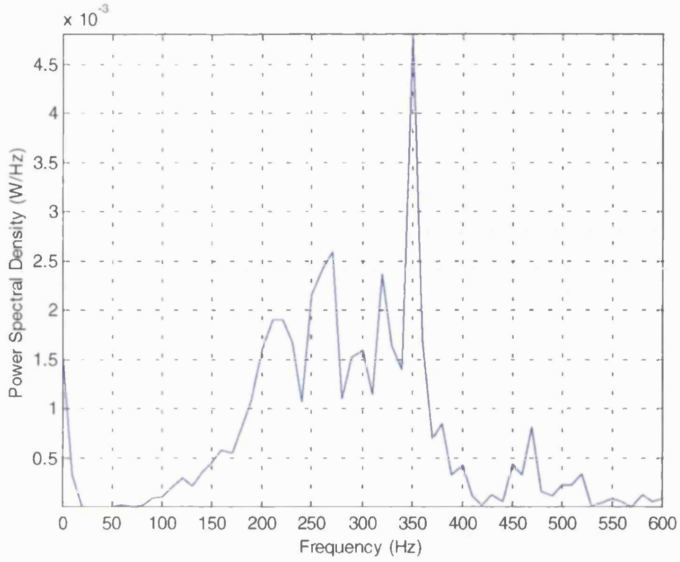


Figure 7.4 - Power spectrum of an interferogram from group 10 acquired with the translation stage

Fig. 7.5 shows a typical spectrum from an interferogram calculated as described in section 6.2 (spatial interferograms) from fringe images acquired with a CCD camera. 100 Hz bulb flicker is no longer a problem, as the interferogram is obtained spatially. Averaging three lines in the image removed most of the electronic noise, and the need for filtering is not critical, due to the good SNR.

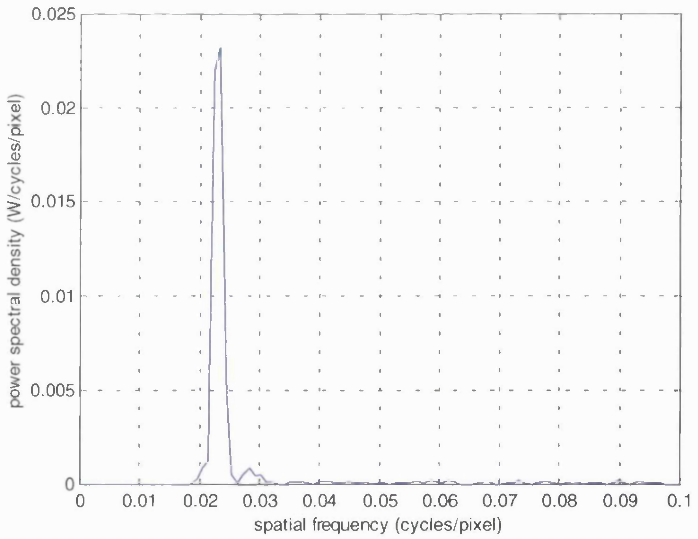


Figure 7.5 - Power spectrum of an interferogram acquired with the CCD camera, in the spatial (non-imaging) mode

In fig. 7.6, a spectrum is displayed that was calculated from temporal interferograms acquired from one pixel of the CCD camera, using the temporal method described in section 6.3. It shows the carrier frequency (highest peak) at approximately 1.5 cycles/micron, in agreement with the expected value, as one period of the interference wave must correspond to its wavelength, in this case 0.633 μm. A noise component can be seen very close to this frequency, with almost the same intensity. The proximity and intensity of this component is an indication of the difficulty in employing the software filtering which is part of the IPSS method using the hardware available for the experiment. The translation stage was the noisiest method of the ones used in the experiments, and the interferograms calculated from the sequences of image frames also captured this noise, represented in fig. 7.6 by the spike to the right of the carrier frequency. The data in this figure was acquired with the He-Ne laser target, in the smallest SCR for which a detection happened. Above that SCR, the noise peak superseded the signal. Due to the proximity between signal and noise, bandpass filtering was not very efficient, leading to the need of averaging, explained in chapter 6.

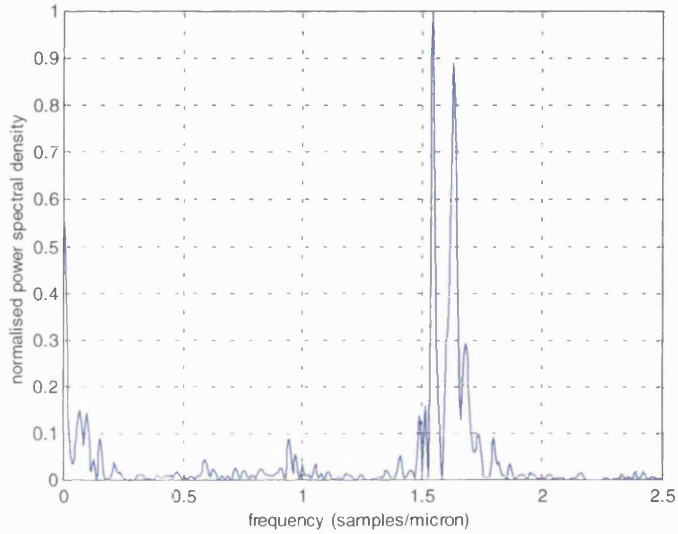


Figure 7.6 - Power spectrum of an interferogram acquired with the CCD camera, in the temporal (imaging) mode

It is important to point out that several different methods of estimating the spectra could have been used, but as the subject of spectral estimation is outside the scope of this thesis, it was decided to use a simpler, but efficient, FFT method. We have investigated the differences between the spectra estimated using this and other popular methods, and found that they are negligible for the accuracy required to design the software filter.

The group ten of interferograms listed in appendix II was used as test waveforms for the design of an algorithm performance assessment program. This group was recorded with a translation stage at 2 Hz, using a He-Ne laser as a target, and a tungsten halogen bulb with an interference filter centred at 632.8 nm as the background. It was chosen in an attempt to check the robustness of the algorithms, as it includes the noisiest interferograms of the experimental data set. A second reason of the choice is that this group contains the results of detection with the He-Ne laser in chapter five, which were used for comparison with the results using other path difference schemes, in chapters five and six, as well as for comparison with the theory in chapter eight.

7.2 Description of the algorithms investigated

In the previous section, it was shown that the signals to be processed by the detection algorithm consisted of a carrier frequency modulated by an envelope and limited in time (or path difference). In the Interferogram Phase Step Technique, the detection is performed through the location of an event marker [Fre93], defined here as a time-localised feature in a signal. Thus, the algorithm to be used to perform the detection of a coherent target in an incoherent background, in the field of view of the interferometer, has to perform two tasks:

i. To detect, in both target-present and target-absent interferograms, the path differences where an event marker occurs; and
ii. To estimate the displacement, in path difference, between these two markers.

The tasks above present an interesting analogy with the classical radar problem, where a target echo has to be detected within a range gate, and its time delay to the transmitted (reference) pulse is then estimated. To accomplish the tasks mentioned, the receiver, understood here as the decision rule implemented by an algorithm, may use several techniques. In a radar receiver, for example, it may just measure the amplitude of the received signal, comparing it to the background noise, or, more efficiently, apply pulse compression techniques, in which a special characteristic (signature) is given to the transmitted pulse, and this characteristic is looked for in the reflected echo, using a decision rule (receiver) that is matched to that signature [Sko90]. The interferograms used in this thesis are quite different from a typical radar waveform, but from the two just mentioned alternatives used in radar, a list of candidate techniques for location of an event marker can be generated:

a) The amplitude envelope of the interferogram may be separated from the fringe carrier, and the path difference for which this envelope reaches a minimum can be found;
b) The instantaneous frequency of the interferogram may be calculated, and the path difference for which this frequency reaches a maximum can be found. In this case, at this position a phase step or transition happens, which is the phenomenon utilised by IPSS;
c) Both amplitude envelope and instantaneous frequency of the interferogram can be used, combining the advantages from both;
d) A template may be calculated from the target-absent interferogram, the correlation between the target-absent and the target-present templates may be computed, and the displacement (in path difference) of the target-present waveform, with relation to the target-absent waveform, which generates a correlation peak can be found; and
e) As the signal under study presents a feature (the phase step), which represents a change in the spectrum of the signal, occurring in a specific moment in time, a

wavelet transform, which can be interpreted as a time-resolved spectral analysis, [Bur98] could be used to locate the phase step.

The methods listed in a) through d) were implemented and are compared in this chapter. Method e), concerning the use of wavelets, was not implemented due to limitations in the time available to perform the research, and is a recommended direction for future work. It has to be pointed out that the list above does not include all possible ways of performing the two tasks needed to process the signals under study. Further algorithms can be created using techniques such as neural networks, or Principal Component Analysis [Kay88], which are outside the scope of this thesis. The three algorithms studied are described in the following sub-sections.

7.2.1. Minimum Amplitude

It was shown in chapter three that the use of an optical interference filter with a steep roll-off characteristic at the entrance to an optical interferometric system gives the interferogram produced from broadband light a sinc-shaped profile, due to the Fourier-transform relationship between self coherence function and power spectrum. It was also shown that the location of the first minimum of the sinc profile in the path difference axis moves to higher values when a coherent or partially coherent emission enters the field of view of the detection system. The displacement of such a minimum can, hence, be used as a discriminant for target detection, and can be measured by determining the location of this minimum for each recorded interferogram. A computer program, listed in appendix I, was implemented in Matlab to locate this minimum. It loads the interferograms, selects a useful region by removing samples at the edges of the interferograms that might contain undesirable artefacts, produced by the experimental apparatus and removes the residual DC component. Although, as mentioned in chapter five, the digital oscilloscope used for signal acquisition was AC coupled, its ground level was manually adjusted, and in most interferograms a residual DC component still existed. After DC removal, the algorithm bandpass filters the interferogram using an elliptic filter, designed to produce a passband ripple of 1 dB and a stopband attenuation of 50 dB. This filtering operation is aimed at removing high-frequency noise and is detailed in the next sub-section, where the IPSS algorithm is described. The analytic signal and the complex envelope [Pro96] of the filtered interferogram are computed, giving its instantaneous amplitude and phase. The calculation of these two signals is explained in chapter three and detailed in the next sub-section. The location of the minimum in the amplitude envelope is then found as the index of the vector formed by the envelope that has the smallest amplitude, and compared with the location of the minimum of the target-absent interferogram, which served as a reference. The difference between these two locations is then calculated. As in chapters five and six, five interferogram pairs were used to generate five differences. Fig. 7.7 presents an example of the use of this algorithm. Plot (a) is the amplitude envelope of the interferogram signal, which is seen in plot (b). On the left of the envelope, the calculated signal displays a non-zero rise time, which is due to the use of the software filter with null initial conditions, but this problem does not jeopardise the calculation of the displacements of the minimum. Another characteristic of the envelope is its smoothness, which limits the accuracy of the location technique. The low derivatives found close to the envelope minimum increase the uncertainty in the location of the event marker mentioned previously.

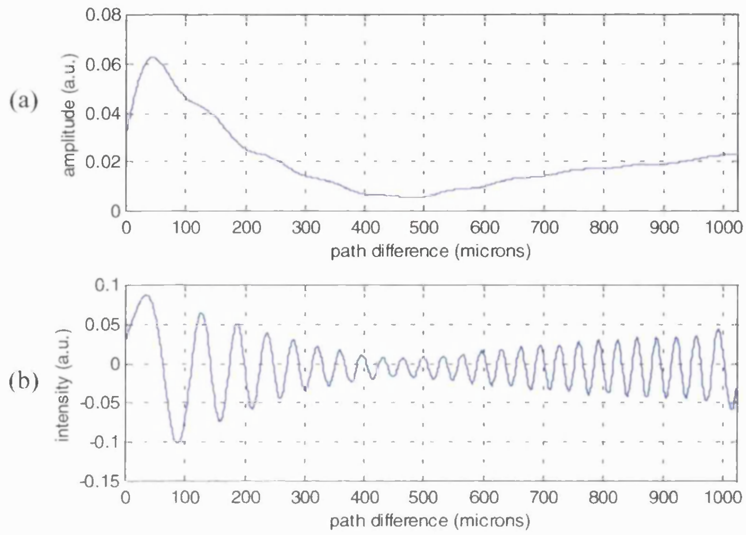


Figure 7.7 – (a) - filtered amplitude (b) - corresponding interferogram.

7.2.2. Maximum Instantaneous Frequency (IPSS)

The IPSS technique as a whole was described in chapter three and, in this sub-section, only the algorithm and its design are presented. The demodulation algorithm (listed in appendix I) comprises three main tasks: bandpass filtering, frequency demodulation, and spike location. As mentioned previously in this chapter, the data available for analysis was corrupted by noise sources inherent to the optical and electronic components employed to perform the experiments. The displacement in the phase step or feature in the path difference axis may be caused by the target, or by unwanted noise sources, which were analysed in the experimental chapters, and are:

I. Fluctuations in target power;

II. Fluctuations in background power;

III. Coupling of power supply mains through the electronics;

IV. Displacement between the spectra of the target and the interference filter, arising from target phase noise, if it is a laser;

V. Noise in the electronics (photodetector, amplifier, oscilloscope);

VI. Drift coming from lack of repeatability in the piezoelectric transducer, or in the translation stage (depending on the experimental arrangement);

VII. Additional modulation of the fringe signal coming from the stepwise motion of the translation stage (not observed when using piezoelectric transducers).

7.2.2.1. Design of the software bandpass filter

In order to remove or reduce the effect of the sources of noise that are spectrally separated from the signal of interest (the fringe carrier and its phase step), a software filter is needed. Its design was based on the following set of requirements:

a) Bandpass, as, from the spectra shown in section 7.1, both low and high frequency components have to be removed, leaving only the fringe signal of interest and the frequency spikes used in the detection process;
b) Linear or zero phase response, as any non-linearity in this response would produce a non-constant group delay, adding different delays to different portions of the interferogram, thus altering the phase step displacement readings;
c) The steepest roll-off possible, as in many situations the 100 Hz caused by mains coupling is very close to the frequencies of interest;
d) Finite filter bandwidth, not too narrowband, as the filter bandwidth must cope with the chirping effect shown in fig. 7.1.

The first choice about the filter concerns the use of finite (FIR) or infinite (IIR) impulse response designs. Although the latter cannot control the phase characteristic, it usually gives lower orders of the filter function than the former to produce the same amplitude response [Ope89], yielding a smaller computational cost. As the required processing is done off-line, that is, not in real time, then it is reasonable to build a filter with zero phase, simplifying the design by considering the amplitude response only. The filter in such a case can only by implemented by a non-causal discrete system [Ope89], that is, a system whose output at $t = 0$ depends on values of the output for $t < 0$, not a problem in an off-line processing application as ours. Thus, an IIR, non-causal design method was chosen. The zero phase filter is obtained [Ope89] by passing the data sequence to be filtered through the IIR filter designed, reverse the output, i.e., reversing the relative order of the samples in the data set, passing the reversed set through the filter again in the same direction, and reversing this second output. The magnitude response is squared, because of the double pass, and the symmetric phase responses cancel. The next step is the choice of the actual filter among the available equations used for analogue filter design, namely Butterworth, Chebyschev I and II, Bessel and Elliptic. These equations were developed for analogue filter design, and represent simple, reliable and useful tools [Ope89]. The option chosen was the Elliptic, because it offers the steepest roll-off and the lowest order possible, at the expense of ripple both in the passband and stopbands. The phase is non-linear, but with the use of the double pass technique it ends up being cancelled.

The elliptic filter just described was applied to the interferograms acquired with the translation stage, comprising group ten of appendix II. As the interferogram carrier frequency was approximately 470 Hz, the design consisted of finding a bandpass elliptic filter with a passband in this region to reject frequencies lower than 400 Hz and higher than 540 Hz, allowing for transition guard bands on each sides. The filter design functions were programmed in Matlab (version 5.3) language through the built in functions *ellipord* and *ellip* to calculate the filter coefficients, with a passband ranging from 440 Hz to 500 Hz, and stopbands below 400 Hz and above 540 Hz. The first function calculates the minimum order required to achieve the design requirements, while the latter implements the IIR filter difference equation [Ope89] using the calculated order. The ripple admitted in the passband was 1 dB and the stopband rejection was set to 50 dB, values that gave the best noise rejection performance, by trial and error. The filter response as calculated by the program is presented in figure 7.8. As expected, the phase response is very irregular, but, as previously mentioned, it is cancelled out by the data sequence double pass.

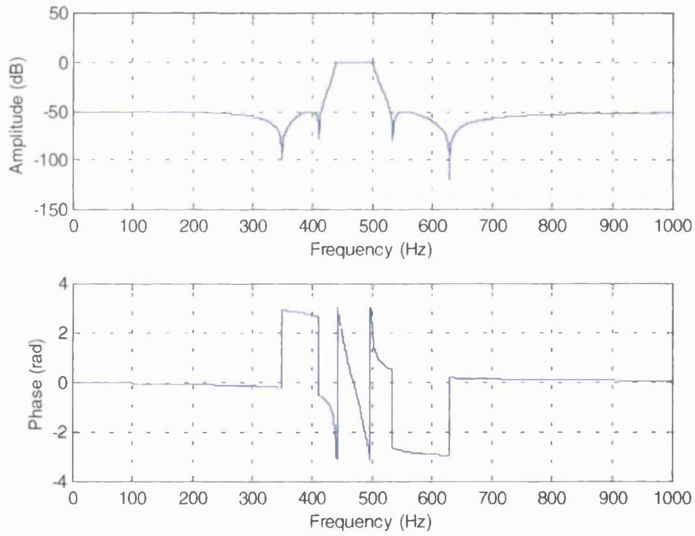


Figure 7.8 - Elliptic filter frequency response amplitude (top) and phase (bottom).

7.2.2.2. Recovery of the instantaneous phase and frequency from the interferograms

The theory concerning the calculation of the instantaneous attributes of a waveform has been included in chapter three. In order to assess such attributes, the calculation of the complex envelope of the waveform is required. In the case of the IPSS algorithm, the signal, after removal of distortion at the edges (beginning and end) of the interferogram (see chapter five) and bandpass filtering, goes through the processing steps indicated in fig. 7.9. The analytic signal of the interferogram is obtained from its Hilbert transform. The result is a modulated narrowband signal, which is then down-converted to baseband through its multiplication by a complex exponential with the frequency of the interferogram fringe carrier. This frequency is calculated using eq. 4.4 ($f = 2.V.\bar{\kappa}$) for the mirror-scanning systems studied in chapter five, depending on the speed employed to vary the path difference (V), and on the central wavelength (or wavenumber k) of the optical interference filter. In the spatial system studied in chapter six, the carrier frequency is simply given by the optical filter central wavenumber (or spatial frequency) instead. The resulting baseband signal is known [Pro96] as the complex envelope, consisting in a sequence of complex numbers with amplitude and phase. The amplitude of the complex envelope is proportional to the real amplitude of the envelope, which was calculated by lowpass filtering in the minimum amplitude method. The phase angle of these complex numbers, calculated using an inverse tangent, is the instantaneous phase of the signal. The next steps in the algorithm are phase unwrapping, to obtain a continuous phase, and differentiation, to obtain the instantaneous frequency. These operations were implemented in Matlab using the built-in function *demod*, which calculates the analytic signal corresponding to the input sequence, multiplies it by a complex exponential at the carrier frequency, calculates the phase angle of the complex result, and differentiates it to find the instantaneous frequency. The calculations performed by the function were checked, and found to be correct, by implementing each of its sub-functions separately, and comparing the results with those from the built-in function, using test interferograms from group one in Appendix II.

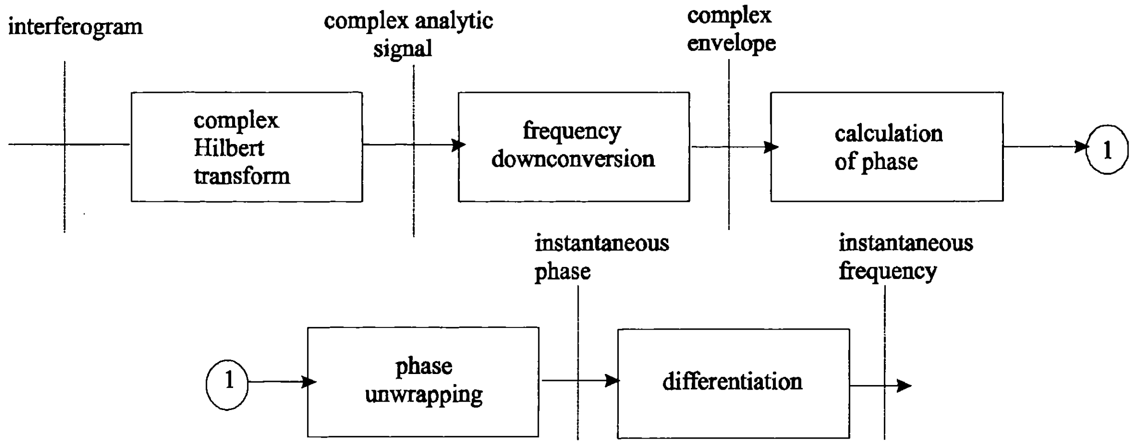


Figure 7.9 - Flowchart of the demodulation module of the IPSS algorithm

Finally, the location of the frequency spikes consists simply in registering the sample number for which the instantaneous frequency sequence coming from the demodulation task reaches its maximum absolute value. The differences in path difference between the locations of the spikes in the target-absent and the target-present waveforms are calculated and recorded for use in the generation of the sensitivity curves displayed in chapters five and six, and in the calculation of the Receiver Operating Characteristic (ROC) curves, later in this chapter.

7.2.2.3. Limitations of the IPSS algorithm

As the main subject studied in this thesis is the detection of coherent or partially coherent targets using the IPSS technique, its limitations are discussed here in detail. Its advantages and possibilities are discussed later in this chapter, when it is compared to the other two approaches studied. In most situations, the algorithm works correctly and the correct frequency spike is detected and located in both target-absent and target-present waveforms. An example of such a situation is given in fig. 7.10. The target-absent waveforms are shown in red, and the target-present waveforms, in blue, both after filtering. In the plot (b), steep phase transitions can be seen in the region where the amplitude (plot (a)) goes through a minimum. In the plot (c), the differentiation of the phase steps forms frequency spikes, which are amplitude-detected to calculate the shifts. Some oscillation can be seen in the demodulated plots (phase and frequency), which are due to a residual error in carrier frequency estimation. In the down-conversion step of the demodulation process, the small difference between the real and estimated carrier frequencies causes the beating seen in the plots. This was found by fine adjustment of the carrier frequency, which eventually removed the oscillations. Even so, this example can be considered as the best case of spike location. The correct spikes, due to the interference filter spectral response, have been detected and are stronger than any other spikes due to noise. However, there were experimental cases among all data groups listed in Appendix II where either the spikes were not detected, or there were stronger spikes detected in clearly wrong positions, which from now on are termed "false spikes". As an example, of the 256 measurements taken in chapter five with a translation stage, 9 gave no spikes, and 11 had stronger spikes from noise.

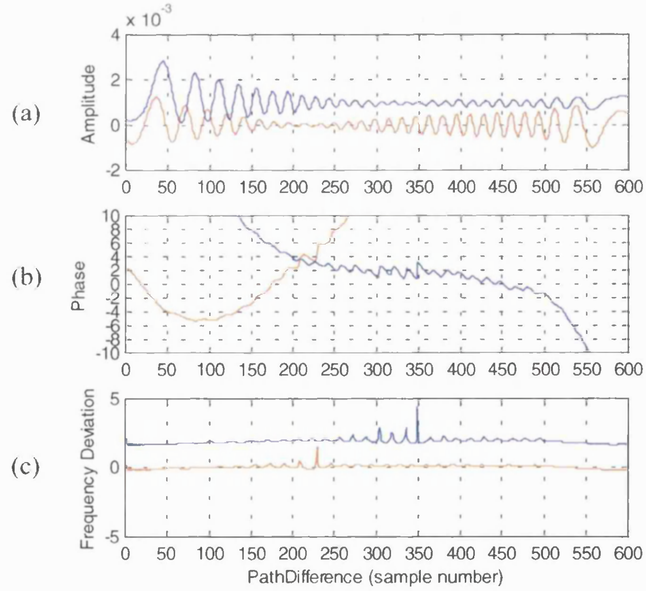


Figure 7.10 - An example of the application of the IPSS algorithm. (a) -Interferogram amplitude. (b) – Instantaneous phase. (c) – Instantaneous frequency. Red - target-absent; blue - target present.

The inability to detect any frequency spikes can be caused by many different factors, related to the interferogram generation and recording process. As mentioned in chapter five, averaging too many waveforms gives a good SNR, but removes the phase step from the interferograms, as these approach a sinusoid. If the bandpass electrical filter, built into the amplifier used to boost the SNR has its cut-off frequencies close to the interferogram carrier frequency, the same effect might happen (but care was taken to avoid this). Another interesting effect was noticed when performing the experiments with the monochromator in chapter five. If the light coming from the target or from the background is poorly collimated, lightwaves with different paths coming from central and extreme rays add within the interferometer [Cha79], and this averaging effect acts the same as when averaging with the oscilloscope, removing the phase steps. A further cause for the lack of detection of the frequency spike occurs when interferograms with high SCRs are measured. When target and background powers are comparable, the interferogram tends to have constant amplitude, depicting no phase transitions. This effect is explained theoretically in chapter eight, and represents an observability cut-off.

The detection of other spikes that are not due to the frequency response of the interference filter can be caused by noise recorded together with the interferogram of interest, due to the causes mentioned at the beginning of this sub-section. In addition to those causes, close to the edges of the interferograms, two effects generate the detection of false spikes: the reversal of the electromechanical device used to vary the path difference, which moves in a cyclic fashion, and the chirping of the interferogram, shown in fig. 7.1. Fig. 7.11 is an example of the former, and fig. 7.12 of the latter. Because of these effects, the developed algorithm included means of removing a number of samples at the edges of the interferogram. Provided that the same number of samples is removed in both target-present and target-absent waveforms, this has no effect on the measured results.

(a)

(b)

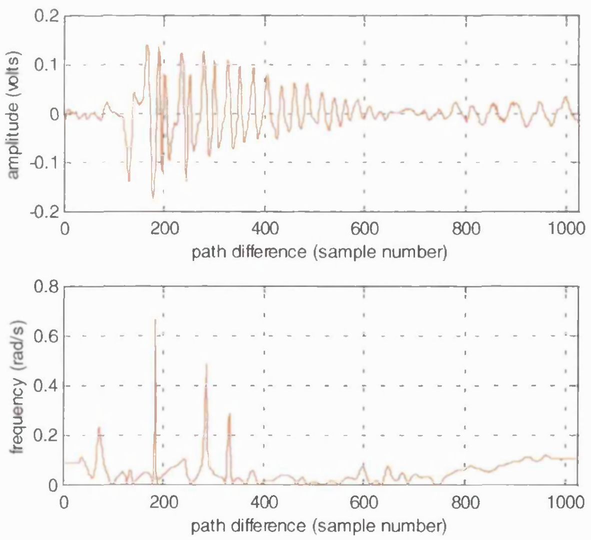


Figure 7.11 - Effect of translation stage reversal on the formation of false spikes. (a) – Interferogram amplitude. (b) – Instantaneous frequency.

Even when false phase steps or transitions are not present in the recorded interferograms, the software filtering process may create false spikes where they do not exist in the original waveform. This may be caused by a poor carrier frequency estimation, which will produce a frequency offset between the real carrier frequency and the passband of the software filter. Due to this problem, the software filter was not used when it was not essential, that is, when the recorded interferogram SNR was good enough for the algorithm to be applied. This happened, for example, with the measurements taken with the fast piezoelectric transducer. Again, if the software filter is used or not in all waveforms of an experiment, for both target-present and target-absent waveforms, no alteration in the result is produced.

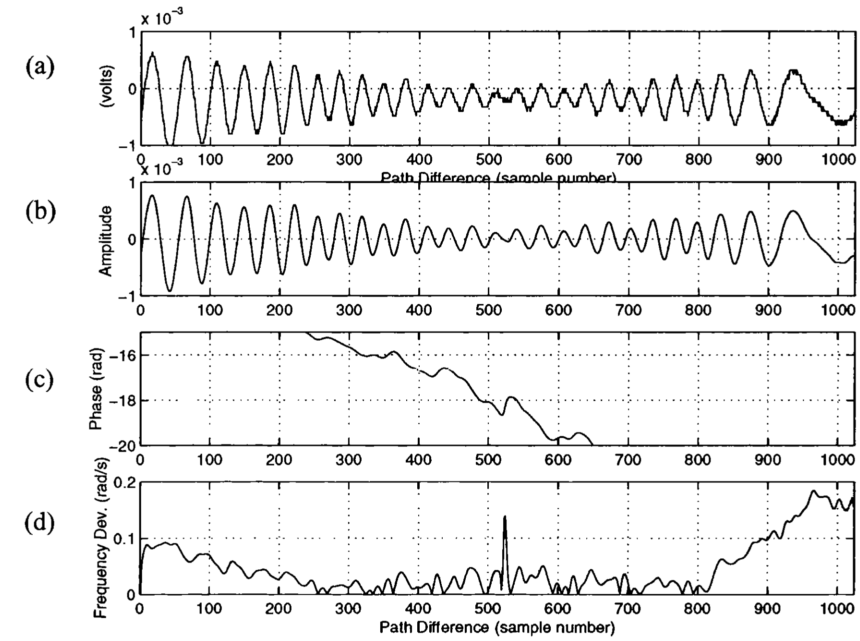


Figure 7.12 - Effect of chirping on the detection of the frequency spike. (a) – Input interferogram amplitude. (b) – Filtered amplitude. (c) – Unwrapped phase. (d) – Instantaneous frequency.

7.2.3. Combination of amplitude envelope and instantaneous frequency (IPSS)

The need for removal of parts of the edges of the interferograms led to the idea of combining the phase and amplitude methods, by using the inverted amplitude as a weighting function to the instantaneous frequency. The region of the interferogram where the amplitude goes through a minimum is, then, given a strong emphasis in the final calculation of the frequency, while other regions are almost discarded. The operations performed by this algorithm can be explained by looking at fig. 7.13. Plot (a) shows the input interferogram. Plot (b) shows the modulus of the complex envelope, which represents the instantaneous amplitude of the interferogram. Plot (c) is calculated from (b) by subtracting the value of each sample from the minimum value found in signal (b), and represents the weighting function (d). This function is multiplied by the instantaneous frequency (a), emphasising the region around the point of minimum amplitude, and rejecting false spikes that might appear out of this region (not in this example), giving the weighted instantaneous frequency in plot (e). It can be seen that plot (e) displays the same result as in plot (b), as, at least in this example, the addition of amplitude to the IPSS algorithm did not improve the ability to find the correct spike. This issue is discussed in detail in the next section.

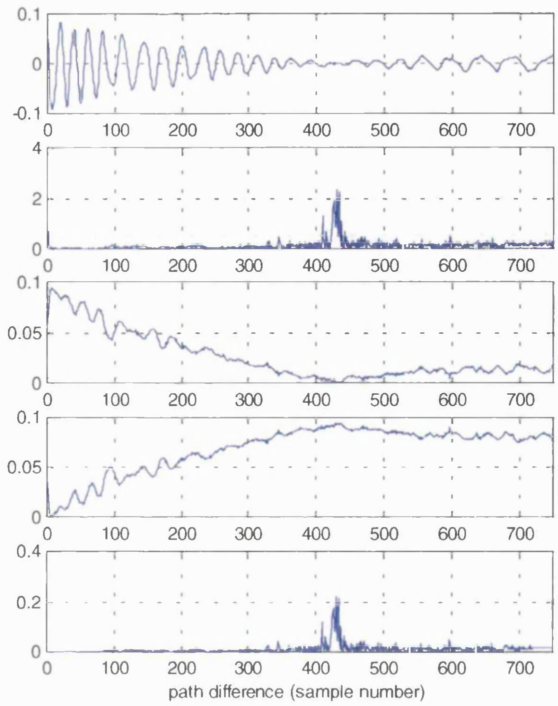


Figure 7.13 - Steps of the combined IPSS/minimum amplitude algorithm. (a) – Input interferogram. (b) – Instantaneous frequency. (c) – Amplitude of complex envelope. (d) – Weighting function. (e) – Amplitude-weighted instantaneous frequency.

7.2.4. Complex correlation

Both target-absent and target-present interferograms possess the same structure, consisting of a carrier modulated by a sinc function close to its minimum, the difference between them being the position of this minimum in the path difference axis. Hence, a possible way of finding the displacement in this position is to move one of the interferograms with relation to the other, and find the displacement for which the structures from both waveforms coincide. This coincidence can be measured by the cross correlation between the two waveforms. In the terminology of signal processing, correlation is, essentially, the same as coherence in optics [Gar92]. The definition of self-coherence was given in chapter three, but here instead of correlating the signal with itself, two different (but not independent) signals are correlated. In the case of two finite, complex signals x and y, with 1024 samples each, the correlation operation evaluates [Ope89]:

$$c_{xy}(m) = \sum_{n=0}^{1023-|m|} x_n y^*_{n+m} \text{ , for } m \geq 0 \qquad (7.1),$$

where m is the correlation delay, and $c_{xy}(m) = c^*_{yx}(-m)$, for m<0.

In order to be effective for use in the optical technique under study in this thesis, an algorithm must measure the displacement in the feature (minimum amplitude, phase step or frequency spike) in the interferogram as accurately as possible. When using correlation to perform this measurement, the performance will depend on the sharpness of the correlation peak, as happens in pattern recognition (see, for example, [Sel02]). Initially we have tried to correlate the whole target-absent waveform with the whole target-present waveform, using their complex envelopes, but this approach did not detect any feature displacement. This was because the displacement that occurs in the phenomenon explored by IPSS is not a simple delay in the whole waveform, but rather its distortion to accommodate the change in the position of the feature, as will be demonstrated in chapter eight. Due to this, instead of using the whole target-absent waveform, we have created a template from it, as follows (see fig. 7.14): for each target-absent / target-present interferogram pair, the analytic signal and complex envelope are calculated as explained in chapter three. The minimum amplitude region of the target-absent interferogram (plot (a)) is found as in the minimum amplitude algorithm (sub-section 7.2.1). A Hanning window [Ope89] with a width of one fifth of the interferogram length is created in the position of minimum amplitude (plot (b)). The interferogram is multiplied by the window, giving the template seen in plot (c), which represents the part of the interferogram close to its minimum. This is then a pattern recognition problem, where the feature with the shape shown in fig. 7.14 (c) has to be recognised within the interferogram. Although the plots in fig. 7.14 display amplitudes, the correlation operates on two complex signals, the template and the target-present complex envelope. Plot (d) shows the modulus of the cross-correlation function. It has twice the number of samples than the interferograms, hence the lag for which the cross-correlation function reaches its maximum value has to be corrected by subtracting the length of the interferogram to give the sought after feature displacement. The performance results of the complex correlation approach are shown in the next section.

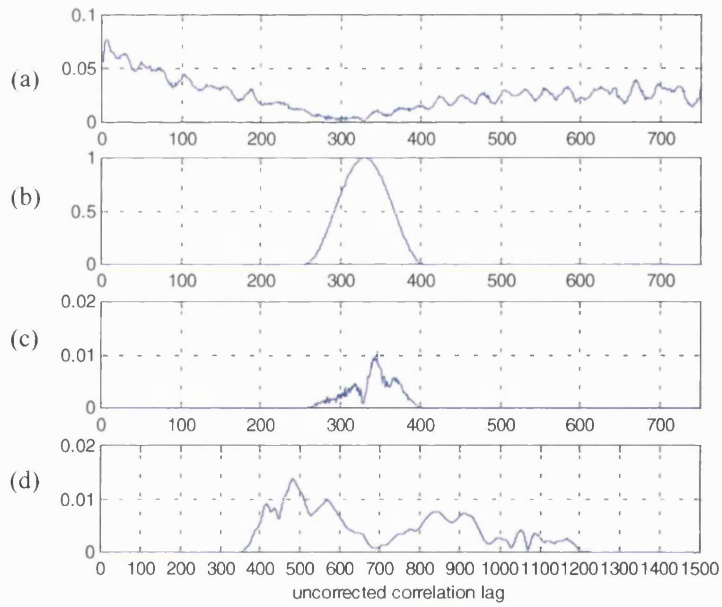


Figure 7.14 - Steps of the complex correlation algorithm. (a) Interferogram complex envelope amplitude. (b) – Hanning window. (c) – Correlation template. (d) – Modulus of cross correlation function. The abscissa axis in plots (a), (b) and (c) is path difference in sample numbers.

7.3. Performance assessment using the Receiver Operating Characteristic (ROC)

The previous section presented four different ways of measuring the displacement, in path difference, of a feature in the coherence profile of the light entering the detection system under study. In order to compare the effectiveness of these approaches, various methods could have been used, such as the simple figure of merit used in the experimental chapters, or computational complexity. As the main concern in this thesis is to characterise the off-line (not real time) detection performance of the proposed approach (IPSS), the method of choice was the construction of Receiver Operating Characteristics (ROC) curves for each algorithm, using the same input data, and the comparison of the curves obtained. The ROC curves are widely used in detection theory [Van68] [McD98], and are a set of curves relating the probability of detection, with the probability of false alarm, having the signal-to-noise (in this thesis, signal-to-clutter) ratio as a parameter.

In order to speed up the evaluation of different algorithms, a computer program (in Matlab) was written to increase the number of interferograms analysed by a single program run. Instead of the target absent-target present pair analysed previously, all the data in the set of measurements was read from disk and processed. In the case of group ten, used as the test set for algorithm comparison, this meant 5 interferogram pairs for each of the 13 signal-to-clutter ratio values, making a total of 130 interferograms. By making the displacement measurement process a sub-routine called whenever each interferogram is read from memory, it was possible to try different schemes much faster, and to assess the effects immediately. This program, its sub-routines and flowcharts are listed in appendix I. The automation of execution of the programs, used to generate faster results, was employed in this chapter only for comparison purposes. The results presented in chapters five and six were generated by applying the IPSS algorithm to one pair of interferograms (target present-target absent) at a time, allowing careful observation of the effects of the different program modules on the data under processing.

In this thesis, so far, the results of the experiments have been presented in graphs where the abscissa was the signal-to-clutter ratio (SCR), defined as the ratio of total powers between target and background, and the ordinate was the shift in the phase step or frequency spike, in μm. Different path difference driving schemes (piezoelectric transducers, translation stage or CCD), as well as the detection of different light sources (laser, LED, monochromator and others) were assessed using the criterion of maximum sensitivity, and this was defined as the minimum SCR for which the one standard deviation error bars centred in the average shifts no longer touched zero. A different criterion could have been established, for example two standard deviation error bars above zero, etc. Thus the minimum detectable SCR depended on the statistical metric of choice.

A more formal approach is to make use of the hypothesis testing theory [Van68][McD98] and to derive the ROC curves applicable to the detection problem under investigation. The algorithm under development is concerned with the decision of whether a target is present or not in the field of view of the detection system. This is done by observing, scan by scan, the position of a particular feature (e.g. a phase step, a frequency spike, or an amplitude minimum) in the path difference axis of an interferogram, and to decide whether it is different from a reference value or not. As five paired samples of interferograms were acquired per SCR, each pair containing one waveform with the target present, and another without it, the problem can be expressed as a hypothesis test on the mean of the differences in feature positions, such as:
H_0: the mean of the differences equals zero, or the two sets (target present and target absent) have the same mean within the test significance level, or, the target is not present;
H_1: the mean of the differences is larger than zero, or the target present set has a larger mean, or, the target is present, and a detection is declared.

In this model the following assumptions were made:

I. As the displacement in feature position in the path difference axis, due to noise, results from the combination of at least the seven factors presented in 7.2, it is assumed that, regardless of the individual probability distributions of the components, there is no dominant component, and the net noise (or the population in statistics jargon) distribution is Gaussian, making use of the Central Limit Theorem. Although it was found in chapter five that there is a systematic error coming from the drift in the piezoelectric transducers and in the translation stage, by acquiring the data in pairs, with approximately 10 seconds in between acquisitions, this drift was considered negligible. Therefore, it was assumed that the noise was random, with no systematic error;

II. Additionally, paired sample testing was used because this reflects the way the data was acquired, in pairs with exactly the same settings, differing only by the presence of a target. It is important to point out that, because the target-present and target-absent waveforms are *not* independent, calculating the statistics (mean and standard deviation) of the target-absent observations and subtracting it from the statistics of the target-present observations is *not* the same as calculating the statistics of the differences between two paired observations [Upt97], which was done in this thesis. This difference was another reason to use paired sample testing; and

III. It was shown in chapter 3 that, without any target-to-filter central wavelength offset, the detection of an emission source produces a positive displacement of the feature in path difference, that is, the position of the feature in the path difference axis moves to a larger value. Due to this, by assumption, the hypothesis test is unilateral, not contemplating the possibility of a negative displacement.

The test statistic to be used in the test is the Student's "t", as the sample size is small (five). It was calculated using [Upt97]:

$$t = \frac{\bar{d}}{s/\sqrt{n}} \qquad (7.2),$$

where $\bar{d}$ is the mean of the differences between the position of the features with target present and target not present; s is the standard deviation of the differences, given by $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2}$, the best unbiased estimate of the population standard deviation, and $n = 5$ (number of samples). The probabilities of detection, (P_D) and of false alarm (P_{FA}) can then be calculated by integration of the t-distribution curves [McD98] corresponding to noise alone, and noise plus signal (displacement), using a selected detection threshold. The distance between the two curves is the value of t given by eq. 7.2 [Upt97]. As both probabilities will depend on this threshold, which is one of the integration limits, it can be varied to produce a plot of P_D vs. P_{FA}, for a specific value of t. Fig. 7.15 shows the threshold, the noise alone and noise plus target probability density curves.

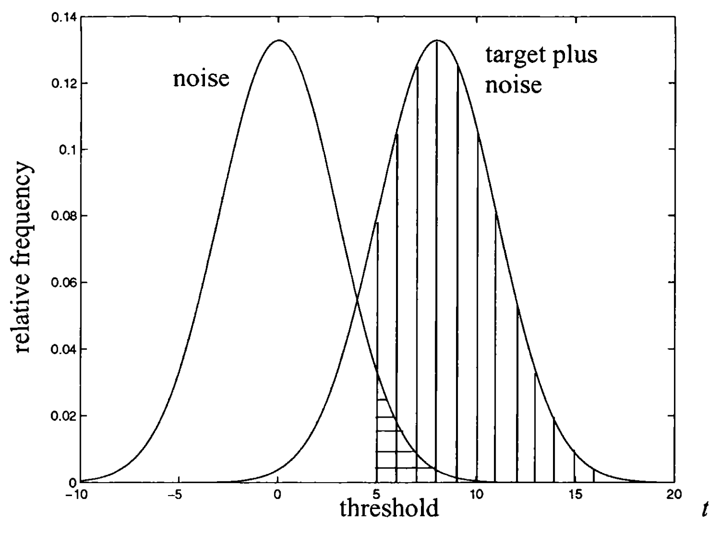


Figure 7.15 – Noise alone (left) and noise plus target (right) probability density functions. The probability of detection is the area marked with vertical lines. The probability of false alarm is the area with horizontal lines.

In a conventional detection system, the value of t is plotted as the signal-to-noise ratio, which has to be larger than one to allow detection with reasonable confidence [Van68]. One of the main advantages of the IPSS technique is that the test statistic t is mapped as the SCR, once for each SCR value there is a displacement in the frequency spike, whose statistic is evaluated by t. For example, a value of t of 1.8 would correspond to an SNR of +2.6 dB (10xlog(1.8)) in an intensity-based detection system, but this same value could be achieved using IPSS at an SCR of –30 dB (calculated using the mean and standard deviation of the phase step shift at an input SCR of –30 dB). It is because of this mapping that high sensitivities can be achieved. Another useful curve can be created using the Neyman-Pearson decision criterion [McD98], largely used in radar and sonar. A value of P_{FA} that is acceptable for the envisaged application is chosen, and a curve of P_D vs. SCR is obtained. The different algorithms described in the previous section can be assessed by choosing the one that gives the largest P_D, for the P_{FA} of choice, and with the smallest SCR.

In order to compare the algorithms presented in the previous section, the method just described was applied to the data contained in group ten of the experimental data listed in appendix II. The reasons of his choice were given in section 7.1. A more comprehensive investigation, consisting of the comparison of the candidate algorithms under all different path difference driving schemes and light sources, was not performed within the scope of this research, and would fully assess, if performed, the robustness of these algorithms. The algorithm comparison program just described loaded the interferograms in the data set and produced as outputs plots of phase step shift vs. SCR (same as in chapter five), P_D vs. P_{FA}, and P_D vs. SCR for a P_{FA} of 0.01, for each of the four candidates. The algorithms were compared using the Neyman-Pearson criterion [Van68] [McD98], as follows:

a) A P_{FA} of 0.01 was defined as acceptable (for example, a P_{FA} of 0.017 is used in [Tar99]), and the minimum SCR that produced a P_D of 50% was recorded for each method and data set. This criterion assessed the sensitivity to small signals;
b) Again for a P_{FA} of 0.01, a figure of merit expressing the average probability of detection in all SCRs was created. This criterion assessed the performance in the whole SCR range.

By applying the above criteria to the four candidate algorithms, we have obtained the data in table 7.1. It can be seen that the addition of amplitude to the IPSS method did not give any improvement in detection performance (the slightly higher figure of average P_D is within the calculation error). This is because the weighting function obtained from the amplitude of the interferogram is much wider in its peak than the frequency spike (as in fig. 7.13), not contributing to reject false spikes relatively close to the correct spike. A suggestion for future work would be the development of narrower weighting functions by reducing the smoothness of the minimum amplitude region (see fig. 7.7). *The best detection performance was that of the IPSS algorithm, as it had the smallest minimum SCR (-14 dB), 5 dB better than the second best, and the highest average P_D (0.2742), more than twice than the second best.* The minimum amplitude algorithm had the second best performance in the average P_D, while the complex correlation was the second best in minimum SCR. IPSS has produced better results than the minimum amplitude method because while IPSS has a displacement resolution limited by the interferogram sampling rate, that is, the smallest detectable shift equals one digital sample, in the minimum amplitude method the resolution is limited by one wavelength of the radiation, once the amplitude of the complex envelope would only be minimum at one of the fringe carrier minima (see fig. 7.14 (a)). The bad performance of cross correlation in average P_D may be a consequence of a template that does not accurately represent the phase step region. A suggestion for future work is the construction of such a template by averaging various interferograms, but compensating for the phase delays between them, usually generated by the drifts in the path difference driving mechanisms shown in chapter five.

Method	Maximum shift (μm)	Minimum SCR (dB) (a)	Average P_D (b)
Minimum amplitude	17	-8	0.1052
IPSS	**15**	**-14**	**0.2742**
IPSS plus amplitude	15	-14	0.2743
Complex correlation	8.5	-9	0.0738

Table 7.1 - Results of the application of the Neyman-Pearson criterion to the four candidate algorithms.

The IPSS algorithm had the best detection performance, using either of the two established criteria, justifying its choice as the main technique studied in this thesis.

Fig.7.16 presents a comparison of the displacements of the event marker (phase step) obtained with three of the candidate approaches (IPSS plus amplitude is not presented because it is the same as IPSS alone), amplitude minimum (red), IPSS (blue) and complex correlation (black), where the abscissa is the SCR measured after optical filtering, for the interferograms of group ten. The three curves display approximately the same shape, with exception of the point at –37 dB of the complex correlation curve, and the point at – 17 dB of the minimum amplitude curve, caused by noise (the corresponding error bars, not shown for clarity, are large). This is the expected shape, as it is similar to that found in the experimental chapters, and is explained in chapter eight. At higher SCRs (-8.5 to -8 dB), the IPSS curve has a peak (at –8.5 dB) followed by a drop (at –8 dB), which represents an observability cut-off, which is detailed in chapter eight, and is not observed in the minimum amplitude curve.

The plots in fig. 7.16 display information solely about the average displacements of the event marker, obtained with the different approaches, not including any information on measurement error. In fig. 7.17, the P_D is shown as a function of the SCR for the three approaches, and is a function of the value of the Student's t parameter (eq. 7.2), which depends on the average and standard deviation of the shift, for each one of the 13 SCRs. The values of minimum SCR listed in table 7.1 can be checked in this figure by reading the SCR where the curves intersect the horizontal line with arrows. The part of the curves below the SCR of – 15 dB can be better visualised in a logarithmic plot. This is done in fig. 7.18, where a change in the slope of the curves is indicated by the circles at –23 dB for the IPSS curve, –18 dB for the minimum amplitude curve, and – 9 dB for the complex correlation. This change indicates that, at that SCR, the algorithm starts to detect the target, and confirms the 5 dB advantage of IPSS in detection sensitivity mentioned earlier in this section.

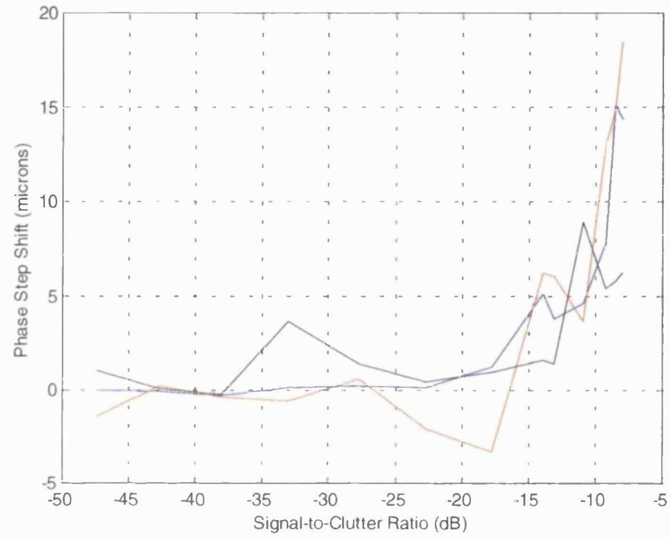


Figure 7.16 - Comparison between the sensitivities obtained with the studied algorithms: blue - IPSS; red - amplitude minimum; and black - complex correlation. The SCR is after filtering.

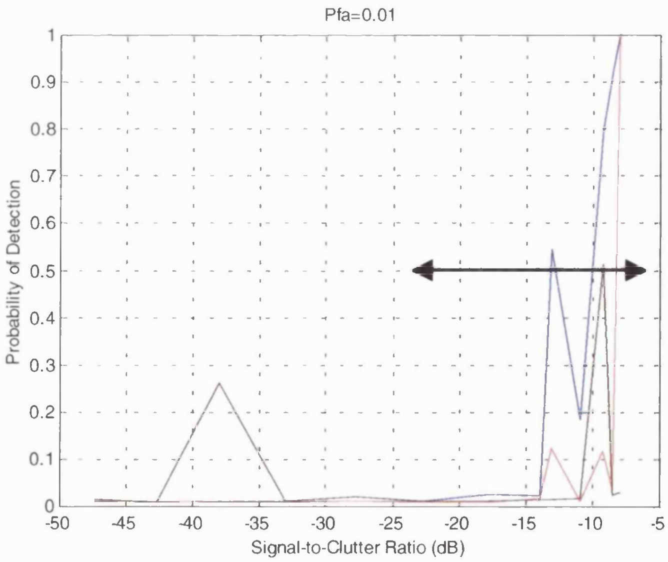


Figure 7.17 - Comparison between the probabilities of detection obtained with the studied algorithms: blue - IPSS; red - amplitude minimum; and black - complex correlation. The SCR is after filtering.

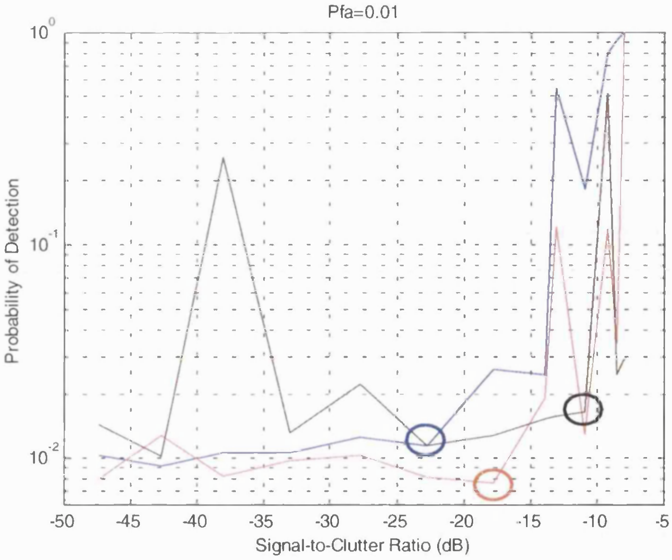


Figure 7.18 – Comparison between the logarithmic probabilities of detection obtained with the studied algorithms: blue - IPSS; red - amplitude minimum; and black - complex correlation. The SCR is after filtering. The circles indicate a change in slope.

The next set of figures presents the set of curves generated by the ROC calculation program for IPSS. Fig. 7.19 is the same shift vs. SCR curve shown in 7.16 in blue, error bars included. The plot closely resembles fig. 5.8, the results obtained with a non-automatic algorithm calculation, apart from three SCR points displayed here and not in chapter five, to allow proper comparison with the theory (in chapter eight). Fig. 7.20 depicts P_D as a function of P_{FA}, with the SCR as a parameter. In this kind of plot, better performance is expressed as a quick rise of the curve to the left of the plot, representing high P_Ds achieved with low P_{FA}s. Between –47 and –14 dB (in the P_{FA} curves), there is little increase in performance, but from –14 to –8 dB the curves quickly shift to the upper left corner of the plot, indicating that detection performance rises sharply after –14 dB, a characteristic due to the corresponding sharp rise of the shift vs. SCR curve (fig. 7.18), which is detailed in chapter eight.

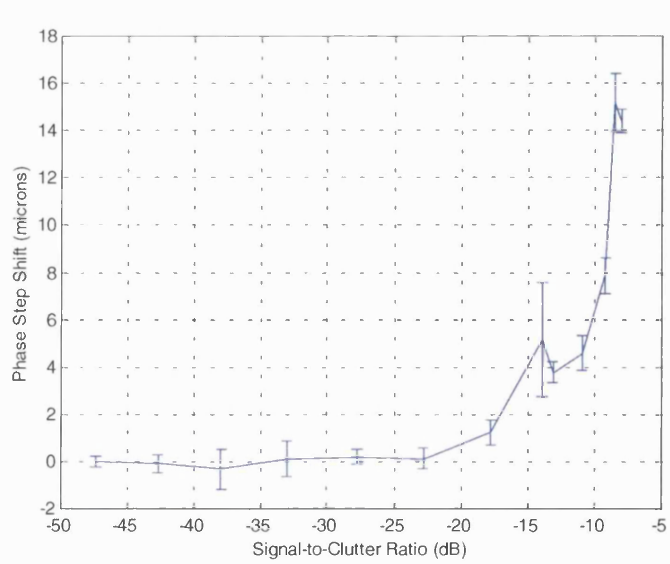


Figure 7.19 - Sensitivity of the IPSS algorithm

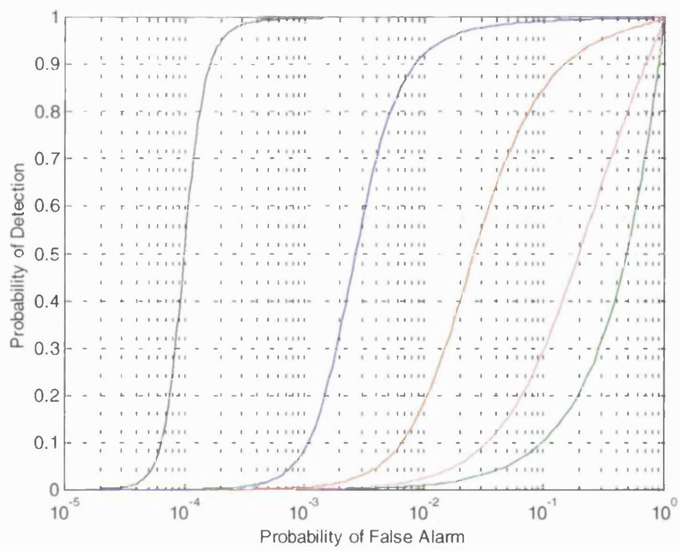


Figure 7.20 – ROC curve P_D as a function of P_{FA}, for the SCRs of –47 dB (green), -14 dB (magenta), -11 dB (red) and –8.5 dB (blue) and –8 dB (black).

Fig. 7.21 has the same data as in fig. 7.20, but in logarithmic scale, allowing better visualisation of the curves. Fig. 7.22 presents the P_D vs. SCR curve for the P_{FA}s of 0.1, 0.01 (also shown in fig. 7.17), and 0.001. The P_D curve resembles the shift curve, with a sharp rise effect happening at –23 dB for a P_{FA} of 0.1, –14 dB for 0.01, and –8.5 dB for the P_{FA} of 0.001. The dips in the curves occur in SCRs where the shift measurements presented larger error, the most prominent being at –14 dB (fig. 7.19). This is because in the region of the curve where the sharp rise starts, the curve has its highest second derivative, and the susceptibility to target and background power fluctuations is increased. Fig. 7.23 presents the same data in logarithmic scale, where a change of slope in all curves can be seen at the SCR of –23 dB.

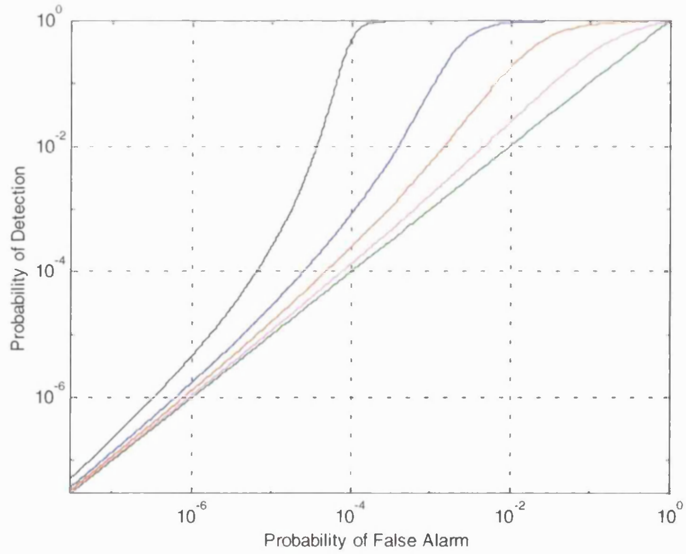


Figure 7.21 - ROC curve logarithmic P_D as a function of P_{FA}, for the SCRs of –47 dB (green), -14 dB (magenta), -11 dB (red) and –8.5 dB (blue) and –8 dB (black).

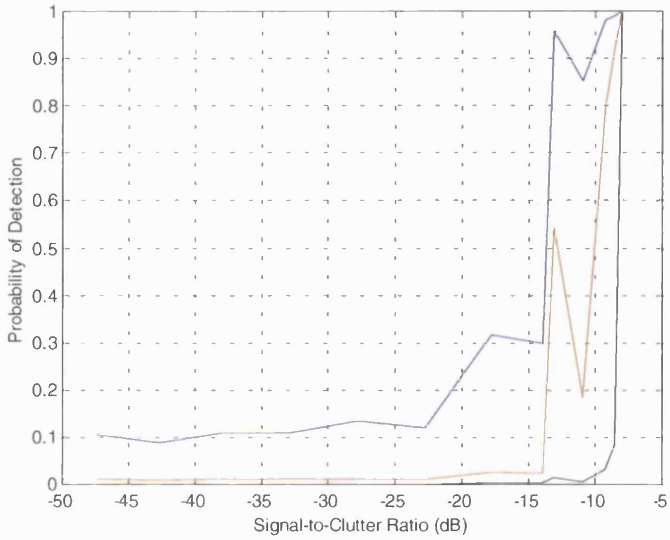


Figure 7.22 - P_D as a function of SCR for the P_{FA}s of 0.1 (blue), 0.01 (red) and 0.001 (black)

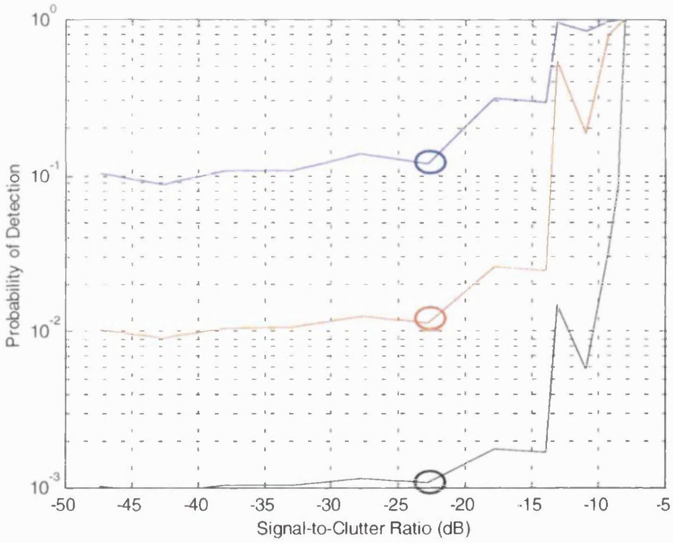


Figure 7.23 - Same data as in fig. 7.22, in logarithmic scale. The circles indicate a change of slope in the curves.

Taking data from fig. 7.19, the sensitivity of the detection system proposed in this thesis, employing the IPSS algorithm, can be characterised more formally as follows: when detecting a He-Ne laser emission, in a background simulated by a tungsten halogen bulb, the system was capable of detecting the target, for example:

1. With an SCR of –8 dB, a P_D of 90% and a P_{FA} of 2.10^{-4};
2. With an SCR of –8 dB, a P_D of 50% and a P_{FA} of 10^{-4};
3. With an SCR of –8.5 dB, a P_D of 60% and a P_{FA} of 3.10^{-4}; and
4. With an SCR of –11 dB, a P_D of 70% and a P_{FA} of 0.04.

Recalling that the SCRs quoted in this chapter were measured after optical filtering, and that this process produces a gain of 18 dB in SCR, for the interference filter centred at 632.8 nm, it is possible to confirm the very high sensitivity achievable using IPSS, with high probability of detection (90%), and low false alarm rates (2.10^{-4}), for very weak targets (-26 dB SCR at the detection system's input).

7.4. Comparison with other detection techniques

In chapter two, several techniques were reviewed, which aim to detect a weak or dim target in a stronger, brighter background. In the literature search performed to write the chapter, it was noticed that there is no agreement on a common figure of merit to characterise the performance of detection systems. Instead, while some authors refer to an SCR gain, some state a clutter rejection figure, and others state the minimum SCR or SNR, both terms appearing to be used as synonyms. In the latter case, the conditions leading to the measurement of a minimum detectable signal are rarely mentioned. In this section, the figures of merit stated in some of the works reviewed in chapter two are compared with the figures obtained using IPSS in the previous section, as well as in chapters five and six. When the papers gave comprehensive information about detection performance (ROC curves), the comparison was done with the results in this chapter. If, however, the authors of previous work just quoted numeric figures without stating the P_Ds and P_{FA}s, the comparison was made with the results from chapter five (experiments with the laser), using the one error bar above zero detection criterion.

Due to the lack of comprehensive information in most papers researched, some assumptions had to be made, in order to allow the intended performance comparison. It was assumed that SNR is the same as SCR. This assumption means that, in the cases studied, the system sensitivity is clutter-limited. Secondly, assuming that no gain or loss is introduced into the signal path, the terms "clutter rejection" and "rejection ratio" are the same as SCR (or SNR) gain. In order to convert a SCR gain to a minimum detectable SCR (MDSCR), it was assumed that a system without signal processing, that is, performing detection on the raw video signal, requires an SCR of +8.2 dB, which was the highest (worst) figure found in the works studied. It is as if not signal processing had been done. Being so, $MDSCR = SCR_{raw} - G_{SCR} = +8.2dB - G_{SCR}$, where SCR_{raw} is the unprocessed SCR, and G_{SCR} is the SCR gain.

In chapter two, several research papers in the detection field were reviewed. In this section, a few examples containing enough information to allow comparison under a minimum set of assumptions were selected for analysis. Concerning the use of ROC curves to characterise performance, Peli and others [Pel97] presented a track-before-detect algorithm used in the detection of sea-skimming missiles in cluttered environments, whose ROC curves presented one point (taken as an example) with P_D=0.85 and P_{FA}=10^{-4}, obtained for a target range in between 3 and 4 km, integrating 10 frames. In fig. 7.20, it can be seen that the performance of Peli's technique is slightly better, but similar to the one obtained with IPSS at an SCR of – 8 dB. This SCR, according to Branlund and others, [Bra96] would correspond, for a sea-skimming missile, to a range in excess of 28 km, and this would have been the range obtained if IPSS had been used. Thus, disregarding the effects of the atmosphere on IPSS sensitivity, and for similar P_Ds and P_{FA}s, **IPSS would detect the target seven times farther** than Peli's technique. Schwarz and others [Sch96] displayed a P_D vs. SCR curve on the detection of military vehicles in terrain backgrounds, of which an example point is P_D=0.8, P_{FA}=10^{-4}, at an SCR of + 6 dB. Again, this performance is very similar to the one obtained with IPSS for an SCR of – 8 dB, representing a **gain of 14 dB for IPSS.**

Some of the papers reviewed did not present ROC curves, but, instead, presented numeric figures related to the minimum detectable signal, as discussed previously. In these cases, as the conditions of the tests were not completely specified, the figures given in the papers were compared to the results obtained in chapter five, using the simplified criterion employed in that chapter (one standard deviation error bar above zero). The works for comparison are listed in table 7.2. When the figure given was a minimum detectable SCR or SNR, it was replicated in the column "MDSCR". When a different figure was given, it was converted using the assumptions previously mentioned.

Reference	Figure-of-Merit	Figure value (dB)	MDSCR (dB)	Conditions
Mon00	SNR gain	52	-43.8	Field data – spatial, temporal and dual-band algorithm
Eis97	Clutter rejection	30	-21.8	Temporal FTS, simulated data
Sch96	SCR gain	21	-12.8	Four bands
Sch96	SCR gain	17	-8.8	Three bands
Tar00	Clutter rejection	16.5	-8.3	Simulated data
Sin98	SCR gain	15.5	-7.3	Off-line processing of field data
Tar99	Minimum SNR	-6.6	-6.6	Simulated target, real background, P_{fa}=0.017, 20 frames, 20 sec delay
Che96	Minimum SNR	-3	-3	Simulated data, laser wavelength measurement
Dia01	Minimum SCR	zero	zero	Laboratory data, 16 frames averaged
Sch96	SCR gain	8	+0.2	Two bands
Bra96	Minimum SNR	+5	+5	Simulated data
Wei95	Minimum SCR	+6	+6	Simulated data, 4 frames, target maximum speed 1 pixel/frame
Wei95	Minimum SCR	+8.2	+8.2	Simulated data, single frame

Table 7.2 – Sensitivity of selected non-coherent approaches, ordered by MDSCR.

In table 7.3, previous work employing coherence is depicted, including the MDSCRs from chapter five of this thesis. In general, the coherence-based methods studied present higher sensitivities (smaller MDSCR) than the non-coherent approaches in table 7.2.

The SCR figures quoted in Drum's PhD thesis [Dru90] were converted from the concentration-path lengths stated in his thesis, 15 ppm.m and 145 ppm.m, using a procedure outlined in appendix IV. Although he claimed a sensitivity of 15 ppm.m for NO_2, which was called "noise equivalent concentration – path length" (NECL) [Dru90], this was the uncertainty in his coherence measurements. We then took the value of 145 ppm.m for comparison, because it is the smallest concentration-path length effectively measured in his work.

The IPSS MDSCR for a laser target was better than all of the figures quoted in table 7.2, but the comparison is not strictly fair as the targets there are partially coherent (missiles and aircraft plumes). When compared to other coherence-based approaches, seen in table 7.3, it is only outperformed by French [Fre88] and Duffy [Duf89]. However, as shown in chapter two, Duffy's claimed results are simply based on the visualisation of a modulation on an oscilloscope screen. French's results [Fre88] cannot be discussed in detail here, as the experimental conditions were not included in his paper. The three figures obtained with coherent targets are possibly equivalent within the errors of the experiments. The result obtained with IPSS for a partially coherent target (filtered white light) represents a fairer comparison with the targets in table 7.2, which are partially coherent as well. The only approach that presented better results was the multiple domain processing by Montgomery and others [Mon00], but, as shown in chapter two, at the expense of a very heavy computational complexity. Concerning coherence-based techniques (table 7.3), the IPSS approach had the best figure among those listed.

Reference	Figure-of-Merit	Figure value (dB)	MDSCR (dB)	Conditions
Fre88	Rejection Ratio	60 dB	-51.8	Conditions not clear; experiment not described.
Duf89	Rejection Ratio	60	-51.8	Experiment in the laboratory, laser target, 12 nm filter, detection criterion is the visualisation of modulation on an oscilloscope screen
IPSS	Minimum SCR	-46.42	-46.42	He-Ne laser target, translation stage
IPSS	Minimum SCR	-31.96	-31.96	Filtered white light target, translation stage
Dru90	Minimum SCR	-27.45	-27.45	Detection of NO_2 in the laboratory, visible band, using the sky as the background, 7.5 nm filter bandwidth, 15 ppm.m
Duf89	Rejection Ratio	30	-21.8	As above, 2 nm filtered white light target, 12 nm filter
Hic88	Minimum SCR	- 20	-20	Experiment in the laboratory, subtraction of two image frames
Dru90	Minimum SCR	-17.6	-17.6	Same as above, 145 ppm.m

Table 7.3 - Sensitivity of coherence-based approaches, ordered by MDSCR. The thick line divides coherent targets (above) from partially coherent targets (below).

As shown in the previous tables and text, IPSS sensitivities outperformed most competing techniques, using either coherent or partially coherent targets. Nevertheless, it is important to stress that the IPSS results quoted were obtained in the laboratory, not including any atmospheric effects, while some (but not many) competing works used field data, with the environmental effects included.

7.5. Conclusion

This chapter presented a comparative study of candidate detection algorithms to be used with the optical technique described in chapter three for detection of coherent and partially coherent targets in incoherent backgrounds. It was shown (in section 7.2) that the IPSS approach has limitations, but those can be circumvented by careful design, alignment and selection of components. The comparison between the candidate algorithms was made using the Receiver Operating Characteristic (ROC) curves as a tool, and the IPSS technique had the best performance among the candidates. The algorithm located the correct frequency spike in 92% of the cases studied, and was capable of detecting a He-Ne laser in a tungsten-halogen bulb background at an SCR of –26 dB, measured at the entrance of the optical system (before filtering), with a probability of detection of 90%, and a probability of false alarm of 2.10^{-4}, considered to be extremely low. This high sensitivity compared favourably with most competing techniques reviewed in chapter two, including coherence measurement [Dru90], a technique with many similarities to IPSS. When the other approaches had a highest sensitivity, either their computational cost was prohibitive, or there was lack of confidence or information about the results claimed by other workers. This sensitivity advantage, along with other advantages shown throughout the thesis and summarised in chapter nine, makes IPSS a recommended solution for the detection of coherent and partially coherent targets in incoherent backgrounds.

Our contributions to the development of the IPSS technique were the use of a frequency spike as the interferogram event marker instead of the phase step invented by French [Fre93], allowing a more accurate location of the marker, and the invention of the first algorithm to implement the IPSS technique, presented in this chapter, which comprised signal processing means of generating the interferogram instantaneous frequencies, of locating the frequency spikes, and of measuring their shifts.

In terms of future work, it might be valuable to further develop IPSS and amplitude methods presented in the chapter to obtain a system with good sensitivity and dynamic range. Other interesting lines of research are the use of complex correlation, with more efficient templates, and the wavelet transform, to find the time in the interferogram where a change in spectrum occurs, due to its phase step.

Chapter 8

Modelling and Performance Prediction

In chapter three, some essentials of interferometry and coherence theories were presented, and the fundamental theory of the IPSS technique was described. The present chapter presents a novel, advanced theory for modelling the displacement in the phase step when a coherent target is detected, which is the kernel of the IPSS technique. The main aim here is to assess the effects of the target to optical filter bandwidth ratio in the system responsivity, for optimisation purposes. A secondary aim is to model the experiments done in chapter five, in order predict the experimental results theoretically and perform a comparison. The chapter starts with the analytical derivation of the self-coherence function of target and background measured by the system's interferometer, only possible under a set of assumptions. The solution is interpreted graphically to yield an assessment of the influence of the target-to-background bandwidth ratio on the system responsivity. The model derived in section 8.1, although giving a good insight of the problem, did not predict the system's detection performance very accurately because of the assumptions. Section 8.2 then presents the computer modelling of the same self-coherence functions for an interference filter with any arbitrary spectral response. The program is applied to calculate the self-coherence function corresponding to the filter used to shape the interferogram in the experiments in chapter five, and a theoretical shift vs. SCR curve is generated. Then the effects of the polarisation of the light sources and optical components in the experiments are studied, and these are introduced into the curves previously obtained. Finally, the theoretical curves for shift vs. SCR, responsivity vs. target-to-background bandwidth ratio and responsivity vs. target to filter central wavelength offset are compared with the experimental results, obtained in chapter five. Good agreement is demonstrated, and the causes of differences are discussed.

In this chapter, the term *power ratio* is used as a synonym for signal to clutter ratio (SCR). The term *sensitivity* is, as in previous chapters, the minimum detectable SCR. The word *responsivity* is also used, being defined as a ratio between the change in phase step shift, and the change in SCR required to cause this change. The sensitivity would, then, depend on the responsivity, and on the amount of system noise.

8.1. Derivation of the system responsivity using a simplified model

In this section we derive an analytical expression for the self-coherence function [Bor99] of the combined target-background radiation. The assumptions made in the course of the derivation are required to simplify the calculations, otherwise an analytical derivation is not possible. In later sections in this chapter, the assumptions are removed, and the model is computer-simulated. In this section, we assume that:

a) The background spectrum is flat across the interference filter bandwidth – This a reasonable approximation, as the solar radiation continuum can be modelled by Planck's blackbody equation [Hud69], with little variation within the optical bandwidth of the filter;

b) The spectral response of the interference filter is a top hat function with unity amplitude, bandwidth $\Delta\kappa$ and centred at κ_0, both expressed in wavenumbers, a good approximation for interference filters, which have steep roll-off characteristics [Hec98];

c) The target is an emission with a Gaussian spectral profile, also centred at κ_0 with a total power P_T and full width half maximum (FWHM), δ, much smaller than $\Delta\kappa$. As explained in section 8.2, this is one of the most common spectral line shapes found in nature or in man-made objects, created by Doppler spreading [Zis93].

As the target and background powers add before being filtered by the detection system, we may write the net power spectrum after narrowband filtering, shown in figure 8.1, as:

$$S(\kappa) = B(\kappa) + T(\kappa) = rect(|\kappa - \kappa_0| / \Delta\kappa)\left\{P_B / \Delta\kappa + \left(P_T.\alpha \big/ \sqrt{\pi}\right)e^{-\alpha^2(\kappa-\kappa_0)^2}\right\} \qquad (8.1),$$

where the first term in the brackets is the background power, with power spectral density of $PB/\Delta k$, so that the background total power through the filter (area under the curve), $B(k)$, equals P_B; the second term in the brackets is the target Gaussian power spectrum, $T(k)$, with a maximum amplitude of $PT.\alpha /\Delta k$, so that the total target power equals P_T [Pis78], and α is a factor inversely proportional to δ. The *rect* function term indicates a rectangular function [Abr72] with unity amplitude, width Δk, centred at k_0.

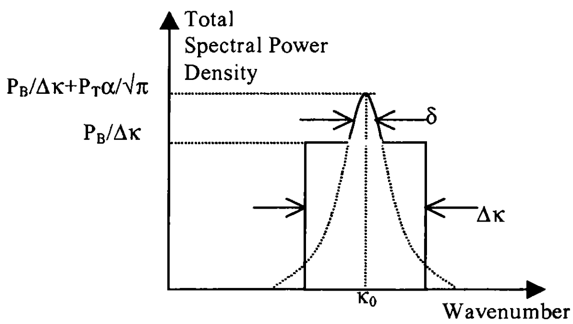


Figure 8.1 - Target and background power spectra

In this chapter, $\kappa=1/\lambda$ is used as the wavenumber, although it is possible to express it as frequency or propagation number, as shown in appendix V. The reason for this choice is that the power spectrum in wavenumbers forms a Fourier transform pair with the self-coherence function given in spatial correlation units (μm), which were used in the recorded interferograms.

As shown previously in chapter three, the Fourier transform (FT) of the self-coherence function $\Gamma(\tau)$ is the power spectrum. Hence, applying an inverse FT to the net power spectrum in eq.8.1 yields:

$$\Gamma(\tau)=\Gamma_B(\tau)+\Gamma_T(\tau) \qquad (8.2),$$

where τ is the path difference and $\Gamma_B(\tau)$ is the self-coherence function of the filtered background and $\Gamma_T(\tau)$ of the target. The calculation of the inverse FT follows. The relationship between α and δ can be derived applying the definition of FWHM to the Gaussian spectrum. It can be shown that $\delta=\sqrt{4\ln 2}/\alpha$. Due to the system linearity, which applies in this case due to the low powers involved in the detection process, one may calculate the Fourier transforms of background and target separately:

$$\Im^{-1}\{B(\kappa)\}=\Gamma_B(\tau)=\frac{P_B}{\Delta\kappa}\int_{\kappa_0-\Delta\kappa/2}^{\kappa_0+\Delta\kappa/2} e^{j.2\pi\tau\kappa}.d\kappa=\frac{P_B}{\Delta\kappa}.\frac{\sin(\pi\tau\Delta\kappa)}{\pi\tau}.e^{j2\pi\kappa_0\tau}=P_B.sinc(\pi\tau\Delta\kappa).e^{j2\pi\kappa_0\tau},$$

a carrier modulated by a sinc function envelope. For the target:

$$\Im^{-1}\{T(\kappa)\}=\Gamma_T(\tau)=\int_{\kappa_0-\Delta\kappa/2}^{\kappa_0+\Delta\kappa/2}\frac{P_T.\alpha}{\sqrt{\pi}}.e^{-\alpha^2(\kappa-\kappa_0)^2}.e^{j2\pi\tau\kappa}.d\kappa$$

A change of variables is required. Let $\kappa_1=\kappa-\kappa_0$. Then:

$$\Gamma_T(\tau)=\int_{-\Delta\kappa/2}^{+\Delta\kappa/2}\frac{P_T.\alpha}{\sqrt{\pi}}.e^{-\alpha^2\kappa_1^2}.e^{j2\pi\tau(\kappa_1+\kappa_0)}.d\kappa_1=\frac{P_T\alpha}{\sqrt{\pi}}.e^{j2\pi\tau\kappa_0}\int_{-\Delta\kappa/2}^{+\Delta\kappa/2}e^{-(\alpha^2\kappa_1^2-j2\pi\tau\kappa_1)}.d\kappa_1$$

To evaluate the integrand it is necessary to manipulate further, adapting a derivation from Kreyszig [Kre88]. For the sake of simplicity, the wavenumber is termed κ rather than κ_1:

Let $v=\alpha\kappa-j\tau\pi/\alpha$ and consider $v^2=(\alpha\kappa-\frac{j\tau.\pi}{\alpha})^2=\alpha^2\kappa^2-j.2\pi\tau\kappa-\frac{\pi^2\tau^2}{\alpha^2}$. Then defining $A=\int\exp(-(\alpha^2\kappa^2-j.2\pi\tau\kappa)).d\kappa=\int\exp(-((\alpha\kappa-\frac{j\tau.\pi}{\alpha})^2+\frac{\pi^2.\tau^2}{\alpha^2})).d\kappa$, where for simplification the integral is represented as indefinite and the notation *exp* is used for exponential. Taking out the exponential that does not depend on κ one obtains:

$A=\exp(-\frac{\pi^2\tau^2}{\alpha^2})\int\exp(-(\alpha\kappa-\frac{j\tau\pi}{\alpha})^2).dk$. But if $v=\alpha\kappa-j\tau\pi/\alpha$, then $d\kappa=dv/\alpha$, thus the integral $I=\int\exp(-(\alpha\kappa-\frac{j\tau\pi}{\alpha})^2).dk$ becomes:

$I = \frac{1}{\alpha} \int_{\frac{-\alpha\Delta\kappa}{2} - \frac{j\tau\pi}{\alpha}}^{\frac{\alpha\Delta\kappa}{2} - \frac{j\tau\pi}{\alpha}} e^{-\upsilon^2} d\upsilon = \frac{1}{\alpha}.J$, where the integration limits were also expressed in terms of v. This is an area under a Gaussian curve. Recalling the definition of the error function $erf(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-u^2}.du$, the integral J can be evaluated as:

$$J = \left[\frac{\sqrt{\pi}}{2}.erf(u) \right]_{\frac{-\alpha\Delta\kappa}{2} - \frac{j\pi\tau}{\alpha}}^{\frac{+\alpha\Delta\kappa}{2} - \frac{j\pi\tau}{\alpha}} = \frac{\sqrt{\pi}}{2}[(erf(+\frac{\alpha\Delta\kappa}{2} - \frac{j\pi\tau}{\alpha}) - (erf(-\frac{\alpha\Delta\kappa}{2} - \frac{j\pi\tau}{\alpha})] \qquad (8.3)$$

In the expression above, an approximation can be made by neglecting the imaginary part of the error function argument (justified in the next paragraph). This can be done if $\tau\pi/\alpha << \alpha\Delta\kappa/2$. Replacing $\alpha = \sqrt{4\ln 2}/\delta$, and analysing the result at a path difference around $\tau = 1/\Delta\kappa$, the region where the phase step in the interferogram occurs (shown later in this section), this condition corresponds to $\delta/\Delta\kappa << 0.664$, or the narrow target case. With this assumption, one has:

$$J = \sqrt{\pi}.[erf(\sqrt{\ln 2}.\frac{\Delta\kappa}{\delta})]$$

Replacing the value of J to calculate I, the value of I to calculate A, the value of A to calculate $\Gamma_T(\tau)$, and putting back the phase exponential due to the carrier, the target self coherence function is:

$$\Gamma_T(\tau) = P_T.e^{j2\pi\tau\kappa_0}.e^{-\frac{(\pi\tau\delta)^2}{4\ln 2}}.[erf(\sqrt{\ln 2}.\frac{\Delta\kappa}{\delta})] \qquad (8.4),$$

where the first exponential is the carrier (fringes), and the other terms are the modulating envelope. The error function accounts for the loss of target energy through the narrowband filter. By neglecting the imaginary part of the argument of the *erf* function in eq. 8.3, which was done in the narrow target case, we have also neglected the apodisation of the Gaussian target spectrum by the filter rectangular function. The effects of this apodisation on the self-coherence function can only be assessed by using a series expansion of the error function of a complex argument [Abr72] and performing a simulation. As in section 8.2 we simulate the whole experiment, including this apodisation, it was decided not to pursue this line of work.

A plot of the error function of a real argument is presented in fig. 8.2 [Pap91]. For $\delta<<\Delta\kappa$, or a narrow target, compared to the filter bandwidth, no attenuation occurs as the error function tends to one. The self-coherence function is a Gaussian. On the other hand, if the argument of the error function is less than 2, or if the target-to-filter bandwidth ratio ($\delta/\Delta\kappa$) is more than approximately 0.416, target power losses start to happen. If the target spectral width is comparable to that of the filter, the narrow target approximation no longer holds, and the solution has to be found numerically, which is done in section 8.2.

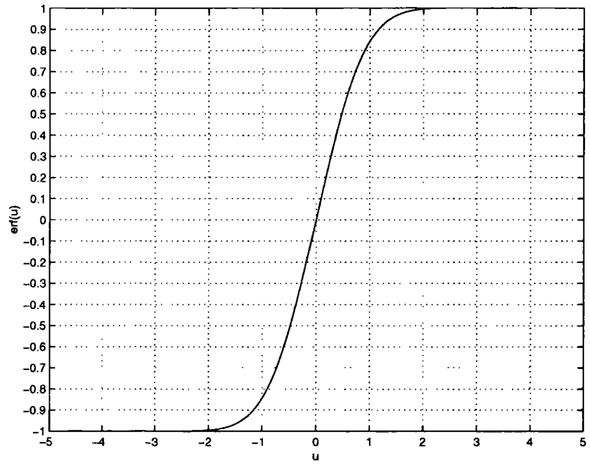


Figure 8.2 - The error function, *erf(u)*

Adding the terms due to the target and to the background one obtains the net self-coherence function of the radiation that reaches the detector:

$$\Gamma(\tau) = (P_B \mathrm{sinc}(\pi\tau.\Delta\kappa) + P_T.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}.erf(\sqrt{\ln 2}\frac{\Delta\kappa}{\delta})).e^{j.2\pi\kappa_0\tau} \qquad (8.5)$$

In order to compare the theoretical and experimental results, we divide eq. 8.5 by P_B, so that the term corresponding to the background has unity power, and the second term becomes '*PR*', standing for the power ratio P_T/P_B. This normalisation allows comparison between various targets and backgrounds with different powers, linearity assumed. The assumption that the behaviour of the system is linear with target and background powers is reasonable if their ratio *(PR)* is kept constant. This is because the change in powers only affects the value along the ordinate magnitude axis of the interferograms, and not the position of the phase step along the path difference (τ) axis. The result is then given by:

$$\Gamma_N(\tau) = (\mathrm{sinc}(\pi\tau.\Delta\kappa) + PR.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}.erf(\sqrt{\ln 2}\frac{\Delta\kappa}{\delta})).e^{j.2\pi\kappa_0\tau} \qquad (8.6),$$

where $\Gamma_N(\tau)$ is the normalised self-coherence function of target and background after filtering.

This equation shows a carrier modulated by an envelope, which is a combination of the influences of the filter and the target spectra. As seen in chapter three, the phase step in the carrier occurs when the envelope function has a zero crossing. Hence, in order to locate this feature, we have to find the roots of the equation. By removing the phase term and setting $\Gamma_N(\tau)$ to zero, we obtain:

$$\mathrm{sinc}(\pi\tau.\Delta\kappa) = -PR.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}.erf(\sqrt{\ln 2}\frac{\Delta\kappa}{\delta}) \qquad (8.7),$$

which is plotted in fig. 8.3. The first root of the equation (or the position of the first phase step) is shown by the circle, in the intersection between the two curves. If no target is present, PR=0 and the first positive phase step position occurs when $\pi\tau.\Delta\kappa = \pi$, or $\tau_{ref}=1/\Delta\kappa$, where the subscript indicates "reference". When a target is present, and PR is increased, the target Gaussian is shifted down in fig. 8.3, and the intersection point moves to the right. The speed of this movement is influenced by the relationship between the width of the spectra from the target (δ) and from the filter ($\Delta\kappa$), which varies the region of the target self-coherence Gaussian that intersects the background sinc curve.

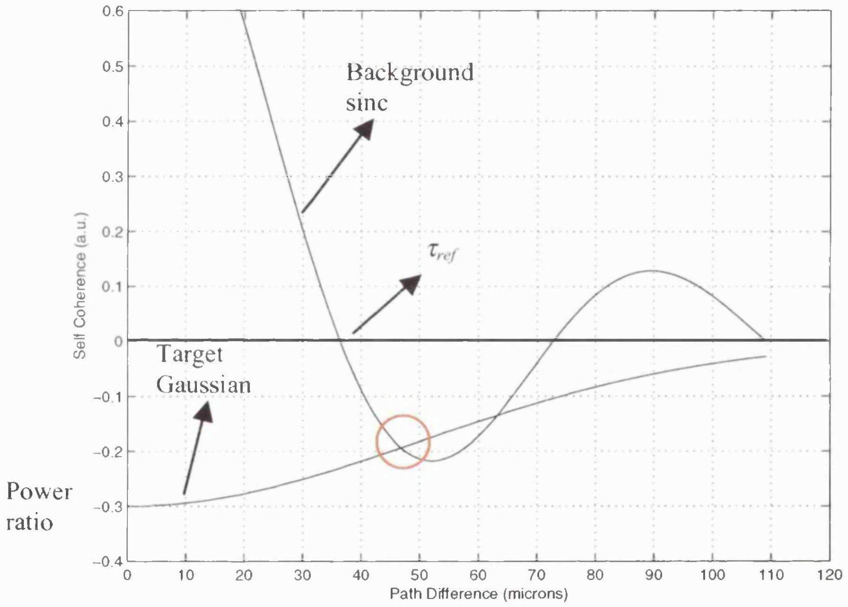


Figure 8.3 - Graphical solution of eq. 8.7, with target bandwidth=3 nm, filter bandwidth=11 nm, power ratio=0.3, and wavelength 633 nm. The circle indicates the first root of eq. 8.7.

As the purpose of this analysis is to assess the effects of the target to background bandwidth ratio on the phase step shift, as the power ratio is increased, three cases of bandwidth ratio are studied, corresponding to three regions of the intersection shown in fig. 8.3 in the target Gaussian curve: the regions close to zero and to infinity, where the slope of the Gaussian is very low, and the region where the Gaussian slope is considerable. In terms of bandwidth ratio, the three regions to be analysed are:

i. $\delta \ll \Delta\kappa$ - This corresponds to a target with a very narrow spectrum (narrow target case), such as a laser, in a much wider filter. In this case, the target has a long coherence length (as shown in chapter three) and the τ_{ref} region is overlapped by a nearly horizontal region in the target Gaussian. The influence of the power ratio increase is to move the target self-coherence (a straight horizontal line in this region) down, pushing the first intersection of the two curves to the right (a higher path difference) and the second intersection to the left (a lower path difference). The maximum measurable phase step shift will correspond to the power ratio where the target line is just touching the background *sinc* function, when only one intersection occurs, and the observability cut-off measured in chapter five is observed;

ii. $\delta \gg \Delta\kappa$ - The target spectrum is much wider than the filter bandwidth, and the filter is blocking most of the energy from the target, as can be seen from the very small argument of the error function. The τ_{ref} region is overlapped again by a nearly horizontal region of the target Gaussian, but with negligible amplitude, leading to very little change in the position of the phase step;

iii. $\delta \cong \Delta\kappa$ - If the two bandwidths are of the same order of magnitude, the intersection between the target and background curves occurs in the region where the target curve has the maximum slope. Although we have derived eq. 8.6 using the narrow target assumption, this is now removed as the error function term in the equation alters only the power ratio (by reducing the target power), and not the influence of the bandwidth ratio in the responsivity, which we want to assess here. By setting the second derivative of the Gaussian spectrum to zero, to find its maximum slope, one obtains $\tau_{max} = \sqrt{2\ln 2}/(\pi\delta)$, the subscript meaning maximum slope. The maximum slope of the target curve increases the speed of the movement of the intersection (or the phase step shift), thus increasing the responsivity. There will be an optimum responsivity point at $\tau_{max} = \tau_{ref}$ or $1/\Delta\kappa = \sqrt{2\ln 2}/\pi\delta$, giving a target to filter optimum bandwidth ratio of:

$$\frac{\delta}{\Delta\kappa} = \frac{\sqrt{2\ln 2}}{\pi} = 0.374781 \tag{8.8}$$

Again in the narrow target case, the observability cut-off mentioned previously in this section is located in the path difference axis at the first negative minimum of the background *sinc* function (in fig. 8.3) at $\tau = 1.43/\Delta\kappa$ (found by setting the derivative of the *sinc* function to zero), from where no intersection between the target and background curves occurs. Thus, the maximum excursion of the phase step position lies within $\tau = 1/\Delta\kappa$ and $\tau = 1.43/\Delta\kappa$. To analyse the dynamic range one has also to consider the width of the target self-coherence Gaussian. In the narrow target situation, the target curve would be nearly a horizontal line, and the upper limit of the dynamic range would be simply the value of the *sinc* function at the first negative minimum ($\tau = 1.43/\Delta\kappa$), or $2/(3\pi) = 0.212$ (see fig. 8.3), corresponding to a power ratio of -6.73 dB. The range of path difference change equals $0.43/\Delta\kappa$, thus inversely proportional to the filter bandwidth.

Still in the narrow target case, it is possible to find how the phase step shift varies with the power ratio *PR*. Analysing in detail the region where the intersection of the target and filter self-coherence functions occurs, one notices that, for a given phase step path difference τ_s , and as the horizontal line from the target self-coherence function moves down with the increase in *PR*, the infinitesimal variation of τ_s with power ratio, $d\tau_s$, is governed by the slope of the *sinc* function at that point, as depicted in fig. 8.4. The power ratio can then be expressed as a function of τ_s using the shape of the *sinc* function. In the region of the sinc curve in fig. 8.3 between $\tau = 1/\Delta\kappa$ and $\tau = 1.43/\Delta\kappa$, one has $PR = \mathrm{sinc}(\pi.\tau_s.\Delta\kappa)$. Unfortunately, the *sinc* function is not reversible, and as our interest is how τ_s varies as a function of PR, the *sinc* function in the interval PR=[0, 2/3π] and τ_s =[1/Δκ, 1.43/Δκ] is plotted in fig. 8.5 with the axis reversed. The responsivity $\partial\tau_s / \partial PR$, defined here as the amount of phase step shift in μm, obtainable from an increase in power ratio in dB, is thus given by an 'inverse *sinc*' function, with no analytic expression, plotted in figure 8.5. In plot a), it can be seen from the line drawn on the plot that the change in τ_s is linear for low power ratios, which is useful in an application such as gas detection, but this does not apply for values higher than approximately 0.1 (-10dB), where a correction based on a look-up table would be required.

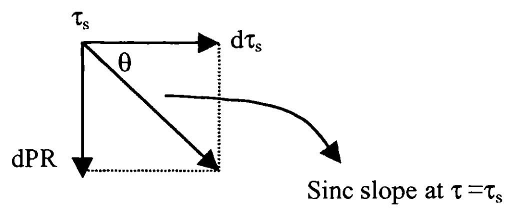


Fig. 8.4 - Graphical interpretation of rate of change of the phase step shift

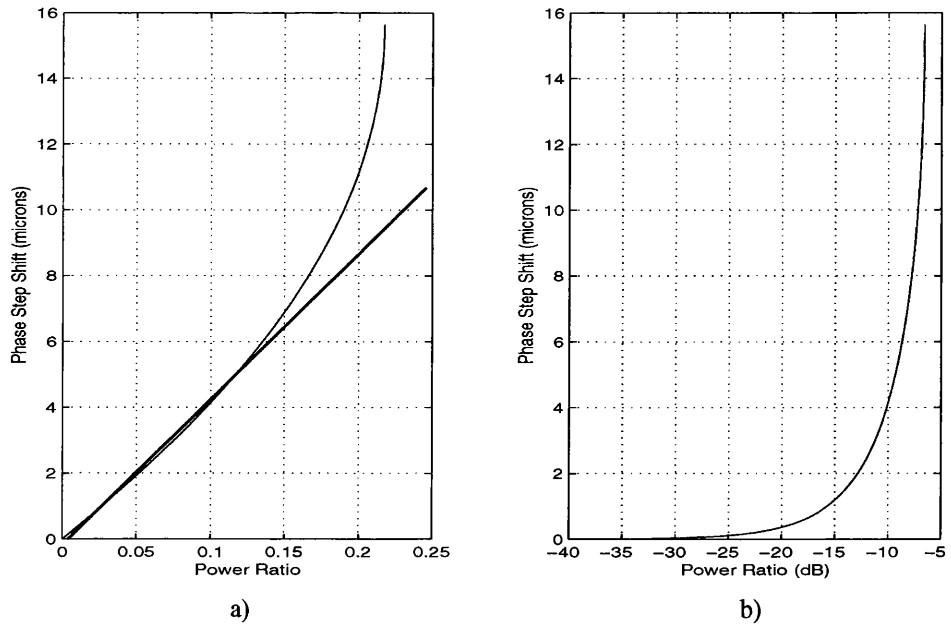


Fig. 8.5 - Variation of the phase step shift with PR of a narrow target, for a linear power ratio in a) and a logarithmic ratio in b). The wavelength is 633 nm and the filter bandwidth is 11 nm.

It is important to assess the question of which type of targets can be considered "narrow targets". By taking the derivative of $\Gamma_{TN}(\tau) = PR.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}.erf(\sqrt{\ln 2}\frac{\Delta\kappa}{\delta})$ (the target normalised self-coherence function, part of eq. 8.6), one obtains:

$$\frac{d\Gamma_{TN}(\tau)}{d\tau} = -PR.erf(\sqrt{\ln 2}\frac{\Delta\kappa}{\delta}).\frac{\pi^2\delta^2}{2\ln 2}.\tau.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}$$

In order to have a narrow target condition, the region of the target Gaussian that intersects the filter *sinc* function must be nearly horizontal, or have a very small slope or derivative. If one calculates the maximum value of this derivative with respect to the product $\delta.\tau$, it can be said that this condition is true whenever the derivative for every path difference is much smaller than this maximum value. By setting τ in the product above to be the maximum value of the interval of interest, or $1.43/\Delta\kappa$ (worst case), one obtains the condition for a narrow target:

$$\frac{\delta}{\Delta\kappa} \leq 0.016 \qquad (8.9)$$

When considering targets with widths comparable to that of the filter, one has also to include the influence of the target self-coherence Gaussian slope in the region where it intersects the filter *sinc* function. As shown earlier in this section, for bandwidth ratios higher than 0.416 the loss of target power in the filter starts to happen, hence one may investigate the region in between 0.016 and 0.416 looking for a change in the responsivity $\partial\tau_s / \partial PR$. As previously calculated, the range of possible path differences ($\Delta\tau_s$=[$1/\Delta\kappa$, $1.43/\Delta\kappa$]) is solely given by the filter bandwidth; however, the range of possible power ratios (*ΔPR*) does vary with the target to filter bandwidth ratio, as different regions of the Gaussian will intersect the filter *sinc* function. When this intersection occurs at a point where the Gaussian has a non-horizontal slope, the power ratio is not limited to $2/(3\pi)$, the value of the *sinc* function at its first negative minimum. This is illustrated in fig. 8.3, where the power ratio (labelled in the figure) is given by the value of the Gaussian for zero path difference, or 0.3 in that case. The responsivity was defined as $\partial\tau_s / \partial PR$, and if we approximate it as $\Delta\tau_s/\Delta PR$, we have that the responsivity is inversely proportional to the range of power ratios *ΔPR*. Hence, for a given filter with bandwidth $\Delta\kappa$, the best possible target is the one that fulfils the narrow target condition, to keep the range of power ratios limited to $2/(3\pi)$, otherwise this range would be longer, and the responsivity lower.

We may find an approximate shape of the variation of responsivity with bandwidth ratio, by noticing that the maximum achievable power ratio *ΔPR* is the amplitude of the Gaussian for zero path difference, when this Gaussian is touching the filter *sinc* function at one single point, as shown in fig. 8.6. This is because, in this case, the value of the *sinc* function at its first negative minimum ($2/(3\pi)$, labelled (a) in the figure) is added to a part of the magnitude of the target Gaussian (labelled (b)).

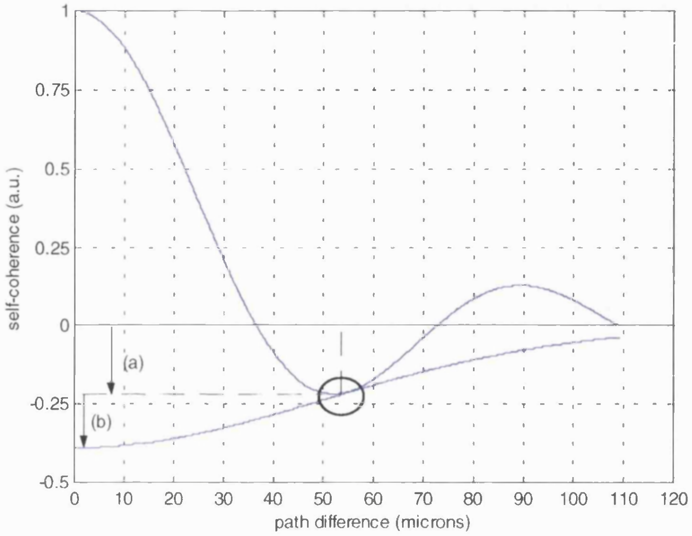


Figure 8.6 – Maximum achievable power ratio ΔPR=(a)+(b). (a) - Value of the *sinc* function at its first negative minimum. (b) - Part of the magnitude of the target Gaussian.

By using the equation of the Gaussian from eq. 8.6 and imposing the existence of the point at which the path difference is $1.43/\Delta\kappa$ and the self-coherence equals $2/(3\pi)$, shown in the figure with a circle, one has:

$\Delta PR.e^{-\frac{\pi^2}{4\ln 2}.(1.43)^2.(\delta/\Delta\kappa)^2} = \frac{2}{3\pi}$. Solving for ΔPR and substituting below gives:

$$\frac{\partial\tau_s}{\partial PR} \approx \frac{\Delta\tau_s}{\Delta PR} = \frac{0.43/\Delta\kappa}{\frac{2}{3\pi}.e^{\frac{\pi^2}{4\ln 2}.(1.43)^2.(\frac{\delta}{\Delta\kappa})^2}} = \frac{0.645.\pi}{\Delta\kappa}.e^{-\frac{\pi^2}{4\ln 2}.(1.43)^2(\frac{\delta}{\Delta\kappa})^2} \qquad (8.10)$$

We can now apply an optimisation technique to eq. 8.10, which shows the responsivity (from now on termed R) as a function of two variables δ and $\Delta\kappa$. The extremes in R=f(δ,$\Delta\kappa$) can be found setting the total derivative of the function to zero [Pis78]:

$$dR = \frac{\partial f}{\partial\delta}d\delta + \frac{\partial f}{\partial(\Delta\kappa)}d(\Delta\kappa) = 0$$

As δ and $\Delta\kappa$ are independent, this is the same as setting each partial derivative to zero:

$\frac{\partial f}{\partial\delta} = \frac{2.02}{\Delta\kappa}.e^{-7.28(\frac{\delta}{\Delta\kappa})^2}.-2x7.28x(\frac{\delta}{\Delta\kappa}) = 0$, a condition which only happens if $\delta \ll \Delta\kappa$. This confirms that, for a fixed filter ($\Delta\kappa$ constant), the responsivity decreases monotonically as the bandwidth ratio increases. Fig. 8.7 plots it as a function of the bandwidth ratio $\delta/\Delta\kappa$ for a given filter bandwidth.

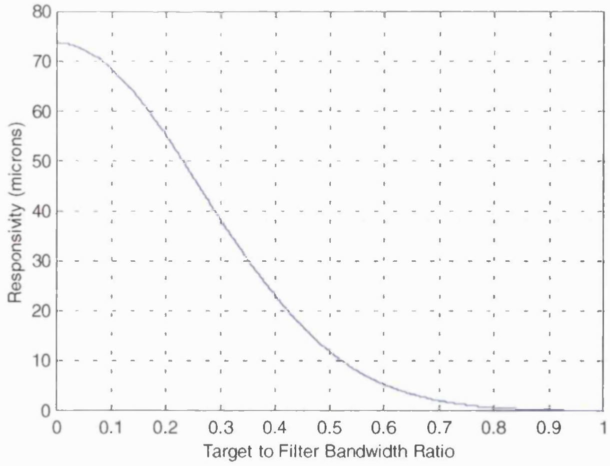


Figure 8.7 - Responsivity $\Delta\tau_s/\Delta PR$ as a function of bandwidth ratio $\delta/\Delta\kappa$ for an 11 nm filter, 633 nm wavelength.

A system designer would be interested in knowing, for an expected target bandwidth δ, what is the filter bandwidth $\Delta\kappa$ that will give maximum responsivity. By taking the partial derivative $\partial f/\partial(\Delta\kappa)$, one obtains:

$$\frac{\delta}{\Delta\kappa} = 0.262 \tag{8.11}$$

This is plotted in fig. 8.8 for a specific target bandwidth. The value in eq. 8.11 is different from the one given by eq. 8.8 because, in that initial analysis, the point of maximum slope of the Gaussian was made to coincide with the starting point of the path difference range, or $\tau=1/\Delta\kappa$, and now, the mean responsivity in the whole range $(\Delta\tau_s/\Delta PR)$ has been used. In the derivation of eq. 8.8, if we have used $1.43/\Delta\kappa = \sqrt{2\ln 2}/\pi\delta$ instead of $1/\Delta\kappa = \sqrt{2\ln 2}/\pi\delta$, we would obtain in eq. 8.8 the same result given by eq. 8.11.

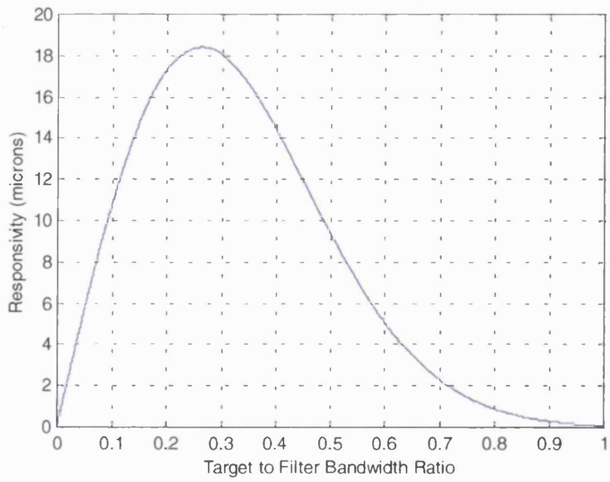


Figure 8.8 - Responsivity variation with bandwidth ratio for a target with FWHM=7 nm.

For design purposes, the optimum ratio to be chosen depends on the application: if one wants to increase the responsivity for low signals or weak targets, then use the value in eq. 8.8; if, otherwise, the aim is to achieve optimum average responsivity throughout the whole range, the value in eq. 8.11 is to be used. The agreement between the plots in figs. 8.7 and 8.8 can be checked by noticing that, for the bandwidth ratio of 0.636 between the 7 nm target and the 11 nm filter, the sensitivities are the same in both plots. As can be seen from the plots, the region where the bandwidth ratio is higher than 0.6 is not of much interest for optimisation purposes, thus the effect of the error function term in eq. 8.6 is not considered in this section, but is simulated in the next.

Figs. 8.9 and 8.10 show two different views of the three-dimensional surface formed by the responsivity R = f(δ,$\Delta\kappa$). The plots in figs. 8.7 and 8.8 are the intersection of this surface with the planes with a filter bandwidth of 11 nm (fig. 8.7) and target bandwidth of 7 nm (fig. 8.8). In order to allow proper display of the surface in fig. 8.9, the minimum target bandwidth in the plot was set to 1 nm, as for a target bandwidth of 0.1 nm, the responsivity reaches 800 microns, using eq. 8.10. Hence, the narrow target region cannot be seen in this plot. The pattern with a point of maximum, obtained with a constant target bandwidth, can also be seen in a lateral view in fig. 8.9. The amplitude of the maximum decreases with the increase in the target's bandwidth.

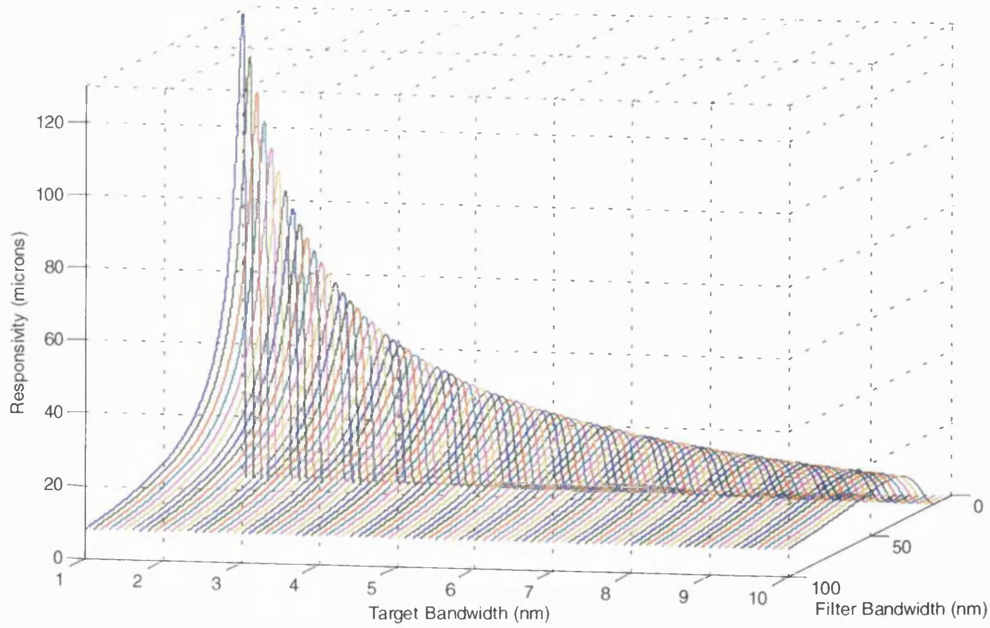


Figure 8.9 - 3D plot of responsivity varying with target bandwidth ranging from 1 to 10nm, and with filter bandwidth, ranging from 1 to 100nm. The wavelength is 633nm.

In fig. 8.10, the curve with a point of maximum is in front view. The responsivity peak that happens for narrow targets is seen in the upper left corner. The region to the lower left where the responsivity is null corresponds to $\delta > \Delta\kappa$ where, as the target is wider than the filter, no change in spectrum or no phase step shift occurs.

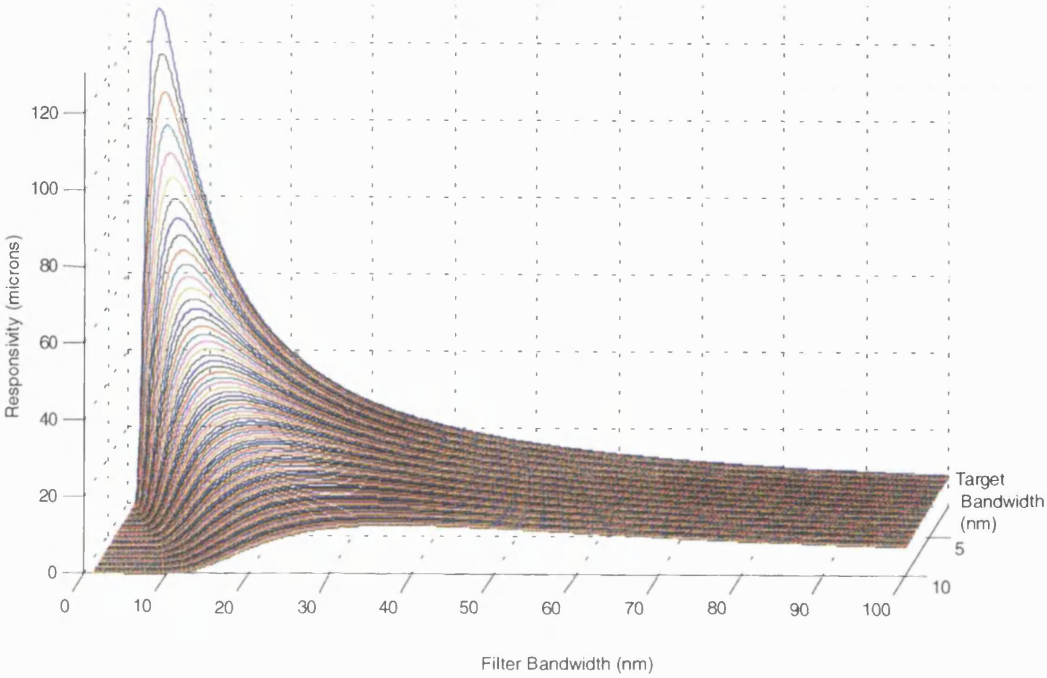


Figure 8.10 - Another 3D view of the same responsivity data displayed in figure 8.9.

The effect of the wavelength in the results presented can be assessed as follows: from eq. 8.10, it is possible to see that the responsivity depends on the bandwidth ratio expressed in wavenumbers. The relationship between the bandwidths in wavenumbers (B_K) and in wavelengths (B_W) is given by [Bor99]:

$$B\kappa = -\frac{B_\lambda}{\lambda^2} \qquad (8.12),$$

where λ is the filter (and target) central wavelength. Applying eq. 8.12 to convert the target and background bandwidths in wavenumbers into bandwidths in wavelengths, and substituting the results into eq. 8.10, one can see that the bandwidth ratio in the exponential term does not depend on the wavelength, as the λ^2 term cancels out. Thus, the optimum bandwidth ratio is independent of wavelength. However, the term that multiplies the exponential in eq. 8.10 is now a function of $\lambda^2/\Delta\lambda$, the filtered background coherence length (using eq. 3.3 from chapter three), where $\Delta\lambda$ is the filtered background bandwidth in wavelengths. Rewriting eq. 8.10 as a function of the filtered background coherence length BCL and the target coherence length TCL (also calculated using eq. 3.3) gives:

$$R = \frac{\partial \tau_s}{\partial PR} \approx \frac{\Delta \tau_s}{\Delta PR} = 0.645\pi.(BCL).e^{-\frac{\pi^2}{4\ln 2}.(1.43)^2(\frac{BCL}{TCL})^2} \qquad (8.13)$$

Hence, the characteristics of the responsivity function given by eq. 8.10, analysed previously, are independent of wavelength, and the responsivity can be expressed either as function of wavenumbers as in eq. 8.10, or of coherence lengths as in eq. 8.13. Combining eq. 8.12 and 3.3 we observe that the bandwidth in wavenumbers and the coherence length are inversely proportional, and the proportionality constant is the square of the wavelength.

In summary, according to the simplified model developed in this section, the detected radiation from target and background can be described by eq. 8.6. The target spectrum is modelled as a Gaussian, the interference filter spectrum as a top-hat function, and the effects of the filter on the spectrum of the target are only considered in terms of loss of target power coupled through the filter. The phase step occurs when the filter self-coherence *sinc* function intersects the target sign-reversed Gaussian (see eq. 8.7), and its shift is given by the displacement of the intersecting point with the increase in the Gaussian amplitude (power ratio, see fig. 8.6). The shape of the phase step path difference vs. power ratio (SCR) curve is an 'inverse *sinc*', and this shape only depends on the filter's self-coherence function (the inverse FT of its spectrum). The effect of the target bandwidth is to pull the upper limit of this curve to higher power ratios, but without increasing the maximum phase step shift, which is limited by the filter bandwidth. The responsivity can be defined and approximated as the ratio between the maximum phase step shift and the maximum power ratio; this approximation was explored for target to filtered background bandwidth ratios up to 0.6, where no target power is lost by filtering. For a given filter, the narrower the target the better, as for a given target, there is an optimum bandwidth ratio, but this figure changes with the range of power ratios, hence depending on the desired dynamic range of the envisaged application. The conclusions above are independent of wavelength.

8.2. Simulation of the system responsivity using a more complete model

The simplified model developed in the previous section allowed the analytical derivation of the shift vs. SCR curves, as well as of the effects of the bandwidth ratio on the responsivity, offering a good insight into the Interferogram Phase Step Shift (IPSS technique). However, the model embeds a set of assumptions which are only valid under specific conditions, limiting the application of the theory to this limited set of conditions, and preventing a broader agreement between theory and experiment. As an example, the approximation given in eqs. 8.10 and 8.13 is very good in the case of narrow targets, but degrades with the increase in the target to background bandwidth ratio. A more complete model was required, which did not present the limitations of the simplified model. In the following sub-sections, the assumptions are listed and a numeric model is created to replace the analytical equations previously used.

8.2.1. Spectral width of the target

The theoretical model presented in the previous section was based on an approximation, which neglected the effects of the apodisation of the target power spectrum by the rectangular response of the optical filter, and was only valid for target-to-filter bandwidth ratios smaller that 0.664. The simplified model, in regard to target bandwidth, is satisfactory to predict the behaviour of the experiments done with the He-Ne laser, but cannot be applied to the experiments with partially coherent targets. The effects of the filter spectral response on the target spectrum, when the target and filter bandwidths are comparable, are the attenuation of target energy, reducing the responsivity, and the creation of sidelobes in the target self coherence function, creating nulls at the same path differences where the background also has nulls. In the limit where the target bandwidth is much wider than the filter bandwidth, no phase step displacement occurs, as an increase in target power does not reduce the effective bandwidth of the target plus background radiation. The computer simulation presented in this section allows the use of any target-to-filter bandwidth ratio, although its limited spectral resolution, given by the length of the FFT used, does not allow accurate simulation of laser targets.

8.2.2. Target spectral shape

In the simplified model, the target spectrum was assumed to be Gaussian. Another spectral line shape with an analytical expression is the Lorentzian. The calculation of the self coherence function of a Lorentzian target is included in appendix VI, but, despite the complex calculations required, the solution did not tend in the narrow target limit to the expected result, an exponential, possibly because of an unidentified mistake in the derivation. Due to this, the filtered Lorentzian formulation was not included in the previous section. Assuming the Lorentzian target is not filtered by the interference filter, it would have an exponential coherence function, due to the FT relationship shown in chapter three. This assumption was used to assess the optimum bandwidth ratio, as in the previous section, obtaining the result of 0.2226, not very different from the one obtained for a Gaussian, 0.262 (eq. 8.11). The shapes of the surfaces plotted in figs. 8.9 and 8.10 (variation of responsivity with bandwidth ratio) were the same in the Lorentzian case, although with less responsivity, as is explained later in this sub-section.

The spectra of the targets used in the experiments in chapters five and six were not measured, as the main aim was to assess their coherence, and were not necessarily Gaussian. The He-Ne laser used has a total bandwidth of 1500 MHz (0.002 nm), as measured by the manufacturer. It has four or five cavity modes (possibly Lorentzian) propagating within the Doppler spreading curve, which has a Gaussian shape [MeG00]. More realistically, the actual lineshape of each cavity mode possibly lies between a Gaussian and a Lorentzian, and is termed a Voigt or mixed profile [Cor88], which can be obtained through the convolution of the two profiles. In the narrow target case (bandwidth ratio smaller than 0.016), the target spectrum does not have any influence on the IPSS system responsivity. As shown in the previous section, if the target spectrum is narrow, it behaves as a spectral line, and its coherence function is a straight line in the region where it intersects the filtered background coherence function, regardless of its spectrum. This is shown in fig 8.11, created using the complex degree of coherence of a top-hat filter (*sinc*), a Gaussian target (Gaussian) and a Lorentzian target, which has an exponential coherence function, due to the FT relationship shown in chapter three. In the right plot, with a power ratio of 0.05 and a bandwidth ratio of 1.8×10^{-4}, simulating the He-Ne laser, both Lorentzian and Gaussian functions are coincident. Hence, in this case, the simplified model of the previous section is valid for any target spectrum.

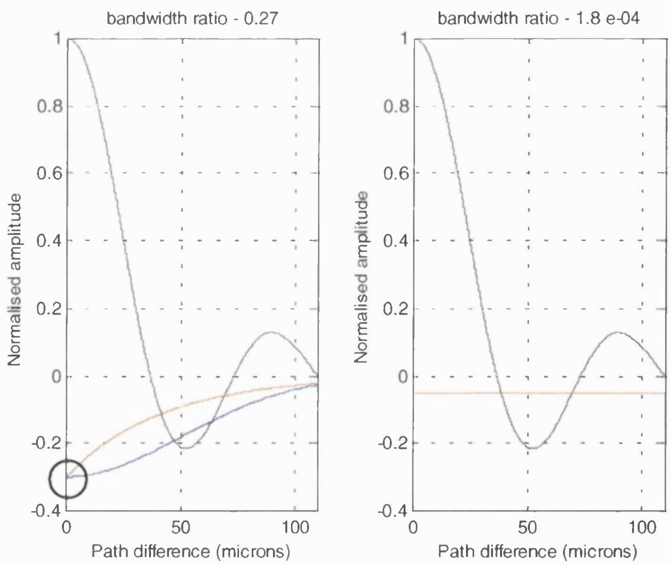


Figure 8.11 – Complex degree of coherence of targets and filtered background for different bandwidth ratios. The black curve is the filtered background, the blue curve is the Gaussian target, and the red curve is the Lorentzian target.

On the other hand, in the left plot of fig. 8.11, with a power ratio of 0.3 (shown by a circle), and a bandwidth ratio of 0.27, the Gaussian and Lorentzian curves are quite different, leading to different intersection points, and different phase step shifts. As, for the same target FWHM, a Gaussian spectrum has a narrower effective spectral width (defined in section 3.1) than a Lorentzian one, its self-coherence function is wider [Goo85] and has a longer coherence length. This fact makes the IPSS approach more sensitive to Gaussian than to Lorentzian targets, due to the longer coherence lengths of the former, and can be confirmed in the plot, where the intersection point of the Gaussian is located at a larger path difference than that of the Lorentzian. This means that, for the same power ratio, a larger path difference is achieved, hence a higher responsivity $(\Delta\tau_v/\Delta PR)$. The simulation model presented in this section calculates the phase step shift for both Gaussian and Lorentzian profiles, and the behaviour of any emission target can be estimated to be within these two.

8.2.3. Filter spectral shape

The simplified model assumed the filter's response to be a top hat function, but the experiment used interference filters with a response which contained a steep, but finite roll-off characteristic, and also strong asymmetry or even a dip in the passband, as shown in fig. 8.12. Interference filters are constructed from multiple, cascaded Fabry-Perot cavities, sometimes employing distributed feedback to achieve higher finesses [MeG99]. As the cavity produces several longitudinal modes, a coating is applied to select the desired modes, which resonates at the desired central wavelength. The combination of these effects usually generates non-uniformities in the passband, as the reflectivity of the cavity mirrors varies with wavelength. These asymmetries, dips and peaks alter the degree of coherence of the filtered background, and the position of the phase step taken as a reference.

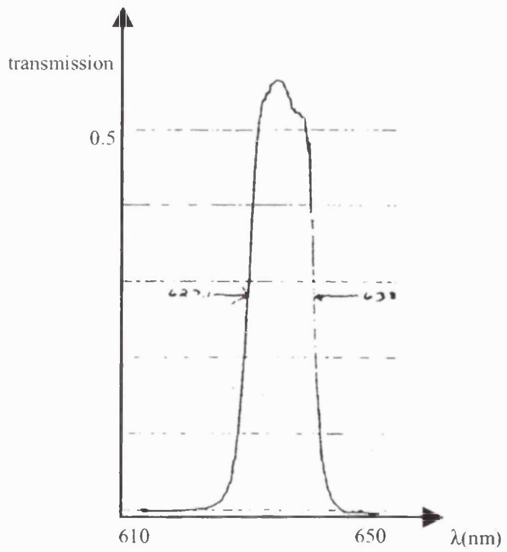


Fig. 8.12 - Spectral transmission of the interference filter used in the laser measurements, with a central wavelength of 632.6 nm, and a FWHM of 11 nm (shown by the arrows)

Fig. 8.12 displays the spectrum from the 11 nm FWHM filter used in the experiments with the laser. It is not only asymmetric, but also there is a slight offset between the target and filter central wavelengths, 632.8 and 632.6 nm respectively. This offset was not considered in the simplified theory, and is discussed in the next sub-section. The simulation model described later can accommodate any filter spectral shape, within the spectral resolution determined by the FFT length. The phase of the filter was not considered in the simulation model due to lack of information from the manufacturer.

8.2.4. Target-to-filter central wavelength offset

Another important issue not addressed by the simplified model is the offset between the target and filter central wavelengths. As the response of an interference filter varies with the angle of incidence [Hec98], in the experiments the interference filters were mounted on a rotation stage, to attempt to eliminate this offset. The method, described in chapter five, did not guarantee this elimination, but only that the integrated power from the target through the interference filter was maximum. As the derivation of the effects of the offset considering the filtering of the target spectrum is very complex, eq. 8.6 from the simplified model was used, neglecting the error function term. If the filter has a central wavenumber κ_0, and the target central wavenumber is κ_1, one has the following self-coherence function:

$$\Gamma_N(\tau) = \mathrm{sinc}(\pi\tau.\Delta\kappa).e^{j2\pi\tau\kappa_0} + PR.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}.e^{j.2\pi\tau\kappa_1}$$, which can be re-written as:

$$\Gamma_N(\tau) = e^{j2\pi\tau\kappa_0}(\mathrm{sinc}(\pi\tau.\Delta\kappa) + PR.e^{-\frac{\pi^2}{4\ln 2}\delta^2\tau^2}.e^{j.2\pi\tau(\kappa_1-\kappa_0)}) \qquad (8.14a)$$

If a Lorentzian target is used instead of a Gaussian, and neglecting target filtering, one has:

$$\Gamma_N(\tau) = e^{j2\pi\tau\kappa_0}(\mathrm{sinc}(\pi\tau.\Delta\kappa) + PR.e^{-\pi\delta\,|\tau|}.e^{j.2\pi\tau(\kappa_1-\kappa_0)}) \qquad (8.14b)$$

Eq. 8.14a differs from eq. 8.6 by the factor $F = e^{j.2\pi\tau(\kappa_1-\kappa_0)}$, representing the difference in carrier frequencies (wavenumbers) between target and filtered background radiation. F is a complex number with unity modulus and a variable phase. If $\kappa_1=\kappa_0$, no offset exists, arg(F)=0, and eq. 8.14a becomes eq. 8.6. At the extremes of the offset interval of interest $\kappa_1-\kappa_0=[-\Delta\kappa/2, +\Delta\kappa/2]$, within the filter bandwidth ($\Delta\kappa$), and around the path difference of interest $\tau=1/\Delta\kappa$, arg(F)=π, or F= -1, meaning that any phase step shift due to an increase in power ratio is negative, that is, to a smaller path difference. At any other path difference, F represents a constant frequency shift between the two terms in eqs. 8.14a and 8.14b, which explains the beating effect measured in the experiments described in chapter five (see fig, 5.17). Fig. 8.13 shows the effect of the offset on the coherence function of a Gaussian or Lorentzian target for a partially coherent target centred at the edge of the filter response ($\kappa_1-\kappa_0=\Delta\kappa/2$). The modulation effect due to the beating brings the inverted coherence functions to positive values. If the power ratio is increased, the amplitude of the blue and red curves increases as well, shifting the intersection points upwards, and causing a phase step shift in the negative direction.

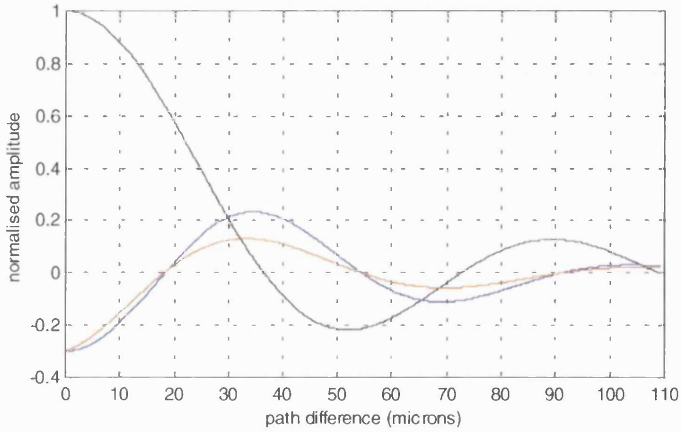


Figure 8.13 - Coherence functions of the filtered background (black), Gaussian target (blue) and Lorentzian target (red), for an offset of half the filter bandwidth

The phase step shift, affected by the envelopes of the two terms in eqs. 8.14a and b, is reduced with an increase in arg(F), and this reduction is symmetric around $\kappa_1=\kappa_0$. This is because the target and background contributions in eqs. 8.14a and b represent a vector summation, where the background has phase zero, and the target has a phase of arg(F), hence the magnitude of the result depends on the cosine of the phase angle arg(F). This is shown in fig. 8.14, which depicts the degradation in responsivity as a function of the offset, for a narrow target (He-Ne laser with a FWHM of 0.002 nm, filter with 11 nm). The figure was plotted using a program in Matlab language that implements eqs. 8.14a and b. Only one curve is shown in the plot because the Gaussian and Lorentzian curves are coincident, as expected. According to the model, offsets larger than 25% of the filter FWHM produce negative phase step shifts. This fact represents an important and novel finding, and was experimentally confirmed in chapter five (see fig. 5.16). An experiment vs. theory comparison is included in the next session.

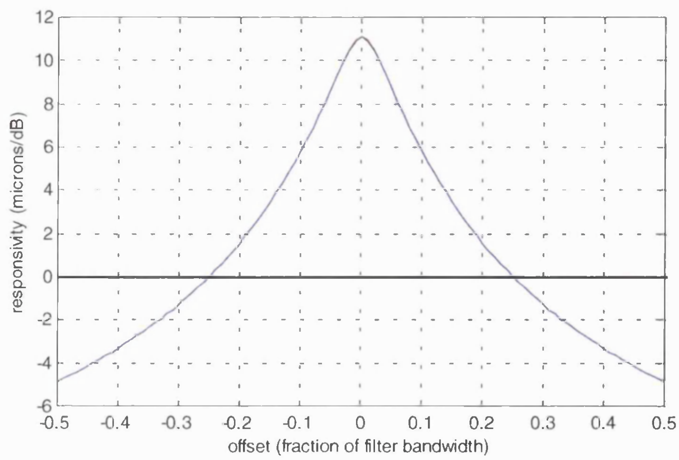


Figure 8.14 - Effect of target to filter offset in the responsivity for a He-Ne laser

Eqs. 8.14a and 8.14b were used to assess the effect of the offset of a partially coherent target with the same FWHM and bandwidth ratio as that employed in chapter five, in the bandwidth ratio optimisation experiment (a monochromator). Fig. 8.15 shows the results of the comparison, which have the same shape as with the coherent target (fig. 8.14), with smaller maximum responsivities. The Gaussian target has a higher responsivity than the Lorentzian, both in the positive and negative directions, but zero responsivity occurs at the same offset (~28%) for the two target types. Therefore, apart from the already known difference in responsivity, the effects of central wavelength offset do not vary with the target spectral shape.

The last remark in this sub-section about the central wavelength offset issue is that this offset causes an asymmetry in the combined filtered background – target spectrum, which, instead of nulls, creates minima in the coherence function. This effect reduces the sharpness of the first amplitude minimum, causing further degradation in the detection performance, as the phase transition happening at this point is no longer a step, rather becoming a smoother phase change.

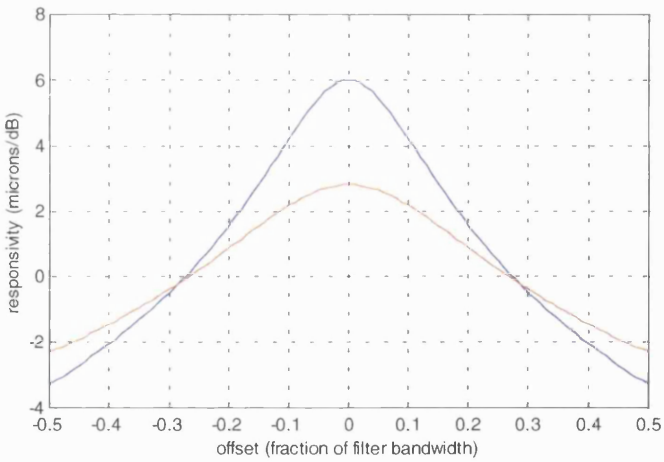


Figure 8.15 - Effect of target to filter offset in the responsivity for the monochromator target used in chapter five. Blue – Gaussian. Red – Lorentzian.

8.2.5. Polarisation

The theory concerning the influence of the polarisation states of the light sources on fringe visibility was presented in chapter three. A set of experiments was performed to assess these polarisation states and their effect on the detection system responsivity. The first experiment was to determine whether the detection system was polarisation-sensitive or not. This was done using a laser without polarisation, a rotating polariser, and the detection system. As shown in chapter three (eq 3.5), the light exiting the interferometer has an incoherent component and an interference term. It was found that the detection system was insensitive to the former, but sensitive to the latter, presenting a selective response with a peak at 72° (clockwise from the vertical), possibly due to the interferometer beamsplitter. A sinusoid was fitted to this measurement, normalised with relation to the maximum fringe visibility (at 72°), and assumed to be the system polarisation response. Any light sources with a different polarisation axis would produce a fringe visibility altered by a polarisation factor given by the cosine of the angle between the source and the system axis. The partially coherent light sources used in the experiments in chapter five were measured to be unpolarised, and the He-Ne laser (also used) was fully polarised. The amplitude of the fringes obtained from the He-Ne laser depended on the orientation of the light source, an effect predicted by eq. 3.9, which describes the influence of polarisation on the self-coherence of fields. Since the phase step shift exploited by IPSS increases with the target to background power ratio, as shown in the previous section, if these fields had different polarisation axes, the fringe visibilities for target and background would be affected differently by the detection system, consequently altering the phase step shift, which is the result of the combination of these visibilities. In the experiment, the laser fringe visibility was measured for different angles of orientation of the laser beam, at every 45 degrees. A sinusoid was fitted to the data, and the angle that produced the maximum visibility response was found to be 36°. Hence, the laser has benefited from a polarisation gain, not seen by the filtered background, for being polarised at an axis relatively close to the system polarisation response axis. Assuming a transmissivity of 50% for the non-polarised light from the tungsten halogen bulb, and given a fringe amplitude reduction factor given by cos(72°-36°)=0.809, the polarisation gain obtained in the experiment in chapter five to detect the He-Ne laser equals, then, $(0.809)^2/0.5=1.309$, where the transmissivity for polarised light was assumed to be the square of the amplitude reduction factor. This gain was introduced into the simulation model through its multiplication by the power ratio, thus increasing the phase step shift.

8.2.6. Other aspects

The white light source used in the experiments is not spectrally flat, as assumed, having a blackbody-like spectrum, with a temperature of 3220 K. Its spectral radiance has a slope of +3.6% within the passband of the 11 nm wide filter, and of +8.7% within the passband of the 36.2 nm wide filter, as calculated using the Planck (blackbody) equation. This will change the spectrum of the filtered light, which will be the product of the background and filter spectra. This effect is included in the simulation presented later.

Finally, the model does not include the effects of noise from the light sources and the detection system. For example, the translation stage used to scan the interferometer's mirror has a non-zero repeatability. This and other effects were discussed in chapters five and seven, and are accounted for in the experiment-theory comparison via the use of error bars in the experimental plots.

8.2.7. The simulation model

A computer program was written in Matlab (version 5.3) to evaluate the phase step shift as a function of power ratio, considering the effects described in sub-sections 8.2.1 to 8.2.6, and is listed in appendix I. The program performs the following tasks, sequentially:

1. Manual input of target data (FWHM) and wavelength;
2. Input of interference filter response, as a table of real numbers between zero and one;
3. Generation of the blackbody curve of the tungsten halogen bulb, using Planck's equation;
4. Calculation of the filtered background spectrum, by multiplying the curve obtained in the previous item by the filter response input in item 2;
5. Conversion of optical bandwidths from wavelengths to wavenumbers;
6. Normalisation of the filtered background, so that its total power (area under the response curve) equals one, and does not alter the power ratio;
7. Normalisation of the filter response, so that its total power (area under the response curve) equals one, and does not alter the power ratio;
8. Calculation of two target-filtered background combined spectra, one for a Gaussian target, and another for a Lorentzian target, using eqs. 8.1 (Gaussian) and A6.1, in Appendix VI (Lorentzian);
9. Calculation of two self coherence functions, by inverse complex Fourier transformation of the spectra calculated in item 8;
10. Calculation of the instantaneous frequencies of the two self coherence functions calculated in item 9;
11. Location of the path difference where the maximum instantaneous frequency occurs, one for each target and each power ratio; and
12. Construction of two phase step shift vs. power ratio curves, one for each target.

The most important experiment modelled for comparison with the theory using the program just described was the detection of the He-Ne laser described in chapter five. Fig. 8.16 shows the target-filtered background combined spectrum, which highlights the issue of spectral resolution as a limitation of the program developed. In order to model the experiment accurately, the ideal spectral resolution would be half the target bandwidth, in the laser case, 0.001 nm, but this sampling frequency was found to be computationally prohibitive, as the PC-computer available for the job did not have enough capacity to run the 2^{24}–long FFT required. However, because the responsivity of the IPSS approach is the same for any target fulfilling the narrow target condition, as shown in section 8.1 (a bandwidth ratio smaller than 0.016, or a bandwidth of 0.176 nm for an 11 nm filter), target and filter spectra were sampled every 0.01 nm, still a lot better than the 0.088 nm required. Another requirement of the program was the ability to predict phase step shifts as small as possible, which depended on the resolution of the result of the inverse Fourier transformation, from the spectrum to the coherence domain. The phase step shift resolution in μm was given by the following equation, adapted from [Cha79]:

$$R = \frac{1/\Delta\kappa}{N} = (\frac{\lambda^2}{\Delta\lambda}).\frac{1}{N} \quad (8.15),$$

where $\Delta\kappa$ is the spectral resolution in wavenumbers, λ is the wavelength, $\Delta\lambda$ is the spectral resolution, and N is the IFFT length, taken as equal to the number of samples of the interferogram. Eq. 8.15 is used in Fourier Transform Spectroscopy, where the length of the interferogram recorded is the reciprocal of the resolution of the spectrum obtained (in wavenumbers) [Cha79]. Eq. 8.15 corresponds, then, to the term $\lambda^2/\Delta\lambda$ (the length of the interferogram converted to wavelengths) divided by the interferogram number of samples, therefore giving the resolution of the interferogram. The IFFT length employed was 262144, which with the spectral resolution of 0.01 nm gave a path difference resolution of 0.15 μm. As this is approximately one quarter of the wavelengths used in the experiments (in the range of 0.6 μm to 0.7 μm, this interferogram resolution was considered adequate for the purpose of calculation of phase step shift vs. power ratio curves.

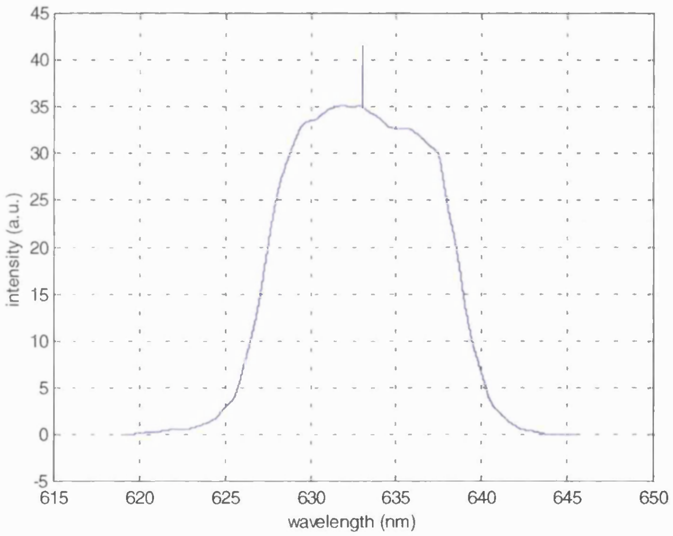


Figure 8.16 - Combined He-Ne laser and filtered background spectrum (11 nm filter FWHM).

The program developed also allowed the simulation of the observability cut-off observed in the experiments and predicted by the simplified theory. Fig. 8.17 presents, in the upper plot, the amplitude (in red) and instantaneous frequency (blue) of the target-background self-coherence function, for a power ratio of −10 dB. The frequency maximum (no longer a spike) is very weak, but still above zero. Although not visible in the plot, the negative peak occurs at the second minimum of the amplitude function. The lower plot presents the same curves, now for a power ratio of −9 dB. The frequency maximum is below zero, hence no longer detectable. The cut-off then occurs at a power ratio between −9 and −10 dB, a value smaller than the one predicted in the simplified theory (−6.73 dB), due to the finite roll-off characteristic of the filter, which makes the frequency spike weaker than in the top hat case. Nevertheless, the experimental value of −8.5 dB was in good agreement with the figures shown here.

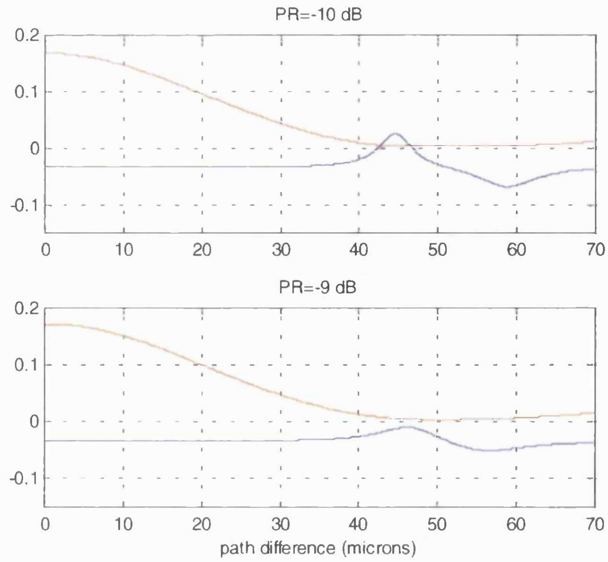


Figure 8.17 - Simulation of the observability cut-off. Self-coherence function amplitude in blue, instantaneous frequency in red.

8.3. Comparison between modelled and experimental results

The simulation model presented in the previous section was used to generate phase step shift vs. power ratio (or SCR) curves for comparison with the experimental results presented in chapter five. Among the different ways of varying the path difference to create interferograms in chapters five and six, only the translation stage offered adequate linearity and phase step shift dynamic range for comparison purposes. An eventual comparison with the experiments done with piezoelectric transducers or with the CCD array would require a far more sophisticated simulation data processing. Fig. 8.18 depicts the curves obtained for the He-Ne laser. The experimental curve, shown in red, was taken from fig. 5.8. As expected, the two simulated curves, shown in blue and obtained considering the target spectrum a Gaussian or a Lorentzian, are coincident. A very good agreement between these curves and the experiment exists up to an SCR (or power ratio) of –13 dB, with the theoretical curves passing within the error bars of the experimental curve. From this SCR upwards, the curves have the same derivative, but there is a horizontal displacement of circa 1 dB. No curve fitting was employed.

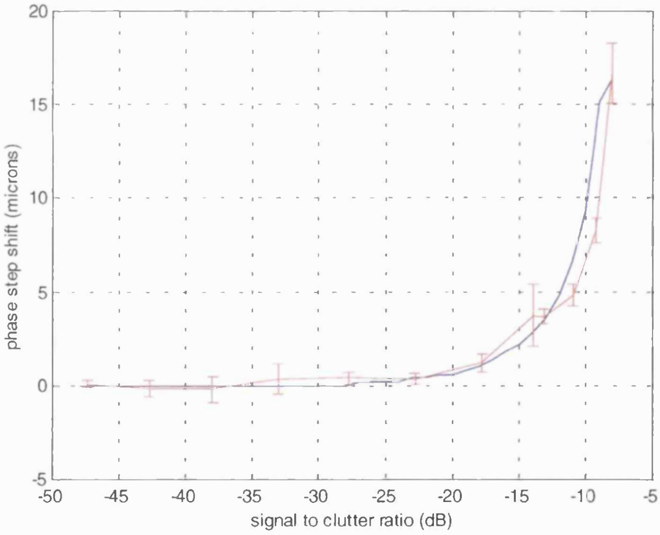


Figure 8.18 - Comparison between simulation model and experiment of detection of a He-Ne laser using the IPSS technique and algorithm.

The power measurements taken during the experiments to calculate the SCR had an uncertainty affected by three main components. Differences in collimation between target and background caused their beams to have different diameters within the detection system, with some light from the background being lost by not reaching the interferometer's mirrors, hence not contributing to the interference which generates the required fringe visibility. Although the photodetector used in the power measurements was filled by both beams at the system entrance, this difference was the dominant component of the uncertainty. Using the areas of the mirrors and of the beams incident at them, this component was calculated to be ± 1.34 dB. Other sources of uncertainty were the resolution of the optical power meter employed to measure power , and the power fluctuation of the light sources during the measurements, which were together assessed by the calculation of an average and standard deviation of three independent power measurements, calculated to be ± 0.61 dB . The calculated overall SCR uncertainty in the He-Ne laser case was ± 1.47 dB, not varying with the SCR. By considering this uncertainty, not shown in fig. 8.17 for better clarity of the plot, the experimental behaviour is fully predicted by the theory to within experimental accuracy. Finally, the simulated curves bend slightly to the right at −8 dB, an effect due to the proximity to the observability cut-off explained previously.

In chapter five, the detection of partially coherent targets was also carried out experimentally. Fig. 8.19 shows a comparison of the results obtained with light from a tungsten halogen bulb filtered by an interference filter as a target. Using such targets, the Gaussian and Lorentzian curves no longer coincide, the Gaussian seen in black, and the Lorentzian in blue, and the former has always produced larger shifts than the latter. This difference is due to the fact that the self coherence function of a Gaussian, another Gaussian, is always wider (has a longer coherence length) than the corresponding function of a Lorentzian with the same FWHM, which is an exponential, as shown in fig. 8.11. In fig. 8.19, the experimental curve is the same as in fig. 5.9, and its large error bars help the agreement between experiment and theory, which is good using a Gaussian target, with one point in seven without agreement, and poor using a Lorentzian target. The target spectral shape is given by the response of the interference filter used to emulate it, which may have spectral structure as the filter seen in fig. 8.12 has, and is neither Gaussian nor Lorentzian, so we do not expect very close agreement. The SCR uncertainty was not measured, because when the experimental data was collected, the beam diameters were not observed. No curve fitting was done in this neither in the following examples in this section.

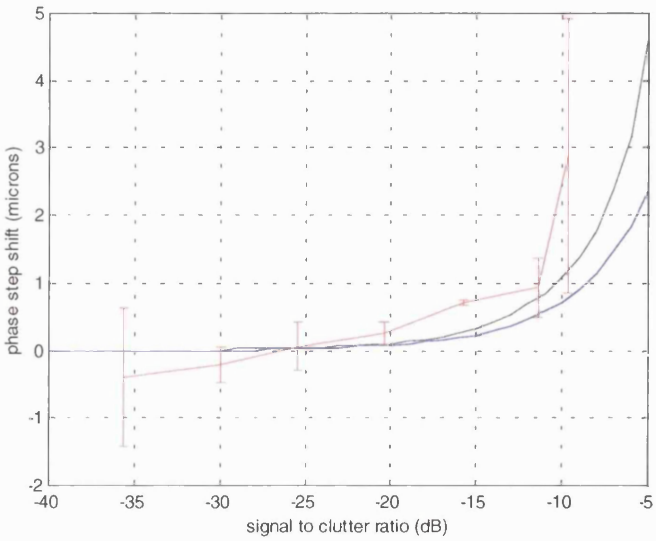


Figure 8.19 - Comparison between simulation model and experiment of detection of a target made with light filtered by an interference filter, using IPSS. Red – experiment. Black – simulation with Gaussian target. Blue – simulation with Lorentzian target.

Fig. 8.20 presents the comparison for the InGaAlP LED measured in chapter five (fig. 5.10). The experimental curve (red) rises faster than both simulated curves, agreeing with the Lorentzian curve (blue) for low SCRs, then with the Gaussian curve (black) for higher SCRs. This change may be due to the increase in current used to achieve different experimental SCRs, leading to different device operation regimes. Even so, the theory is able to predict the experimental result with a reasonable accuracy, and the Gaussian seems to give the best fit.

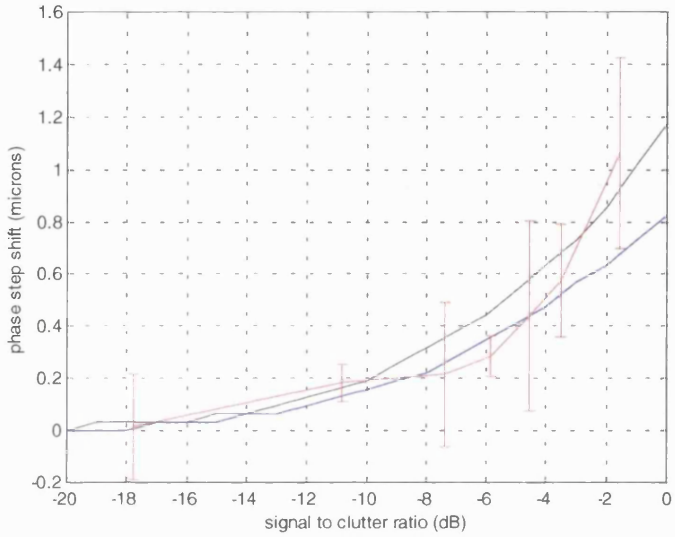


Figure 8.20 – Comparison between simulation model and experiment of detection of a InGaAlP LED using IPSS. Red – experiment. Black – simulation with Gaussian target. Blue – simulation with Lorentzian target.

The last partially coherent target studied in chapter five was the RCLED (fig. 5.11), whose comparison of results is shown in fig. 8.21. The simulated curves were obtained using a target FWHM of 6.5 nm. This value was obtained from spectral measurements done by Oulton and others [Oul01]. The device investigated in this thesis was manufactured by the authors of the paper, thus having similar characteristics with the devices studied there. The device had a central wavelength of 650 nm and the FWHM was measured with a numerical aperture of 0.53, the maximum NA obtainable with the iris-microscope objective arrangement described in chapter four. At this large NA, as explained in chapter four, the device is believed to have a Gaussian lineshape. This is confirmed by the agreement of the experimental curve with the Gaussian curve up to –8 dB, from where the experimental curve has a smaller slope than the Gaussian, and the curves no longer agree. As the experimental SCR was varied by changing the device current, it is believed that this is due to the reduction of 6% that occurs in the coherence length of the RCLED when the current is increased from 5 to 40 mA (see chapter four), making the IPSS approach less sensitive to the RCLED, hence producing shorter phase step shifts. Even with this effect, the experimental curve still lies between the Gaussian and Lorentzian simulated curves.

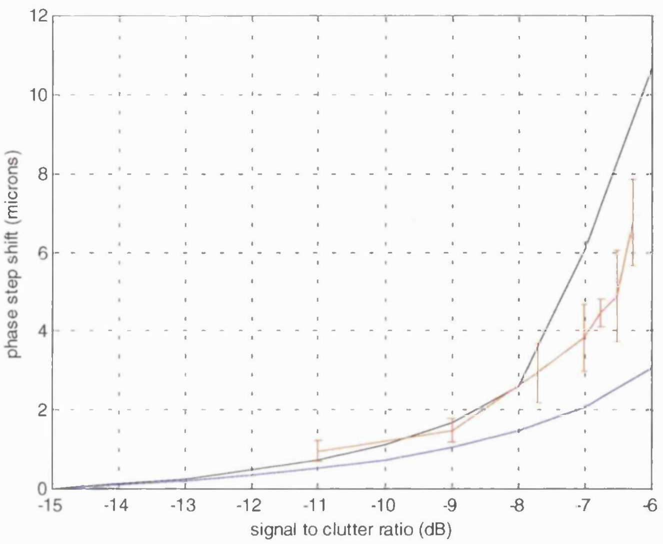


Figure 8.21 - Comparison between simulation model and experiment of detection of an RCLED using IPSS. Red – experiment. Black – simulation with Gaussian target. Blue – simulation with Lorentzian target.

An experiment to investigate the optimum target-to-background bandwidth ratio was performed in chapter five. Its results were copied from fig. 5.14 to fig. 8.22, which also presents a point and error bar obtained with the fast piezoelectric transducer at a bandwidth ratio of 0.3, and a theoretical curve obtained using the simplified model described in section 8.1. It was decided to use this model, valid for narrow targets, due to its ease in producing the required curve. However, its results were compared with those from the extended model described in section 8.2 at three bandwidth ratios, with negligible error. It can be seen that the shape of the theoretical curve is similar to that of fig. 8.8, calculated using the simplified model. The responsivity in μm/dB, for each bandwidth ratio, was obtained by dividing the maximum phase step shift obtained with the model by the difference in dB between the minimum and maximum SCRs, giving the average slope of the shift vs. SCR curve. This method of calculation explains the difference between the curves, once that the minimum theoretical SCR does not account for noise, being smaller than the minimum experimental SCR, yielding a larger denominator and a smaller μm/dB ratio. In spite of this difference, there is agreement between the theoretical curve and the lower limit of the error bars, which very approximately corresponds to no experimental noise. More importantly, the point of maximum responsivity in the experimental data is fairly close in bandwidth ratio to that of the theoretical curve, validating the finding that there is an optimal target to background bandwidth ratio, at about 0.3, which yields maximum responsivity.

Finally, in chapter five, the effects of an offset between target and filter central wavelengths were studied. Fig. 8.23 presents a comparison between the experimental data, copied from fig.5.15, and theoretical results for a Gaussian target, using the same data as in fig. 8.22, converting the maximum shift in μm to a responsivity in μm /dB, as done previously in this section. The experimental data, shown as black squares, was interpolated to produce the black curve. The agreement between the experiment and the theory (red) is good, with a slight asymmetry of the experimental curve, possibly due to an asymmetry in the response of the interference filter used to perform the measurements. It can be seen that, in order to retain the high responsivity IPSS can offer, careful optical filter design has to be done, to minimise this offset. By limiting the offset to +/- 10% of the filter bandwidth, the responsivity is kept at approximately 80% of its maximum value.

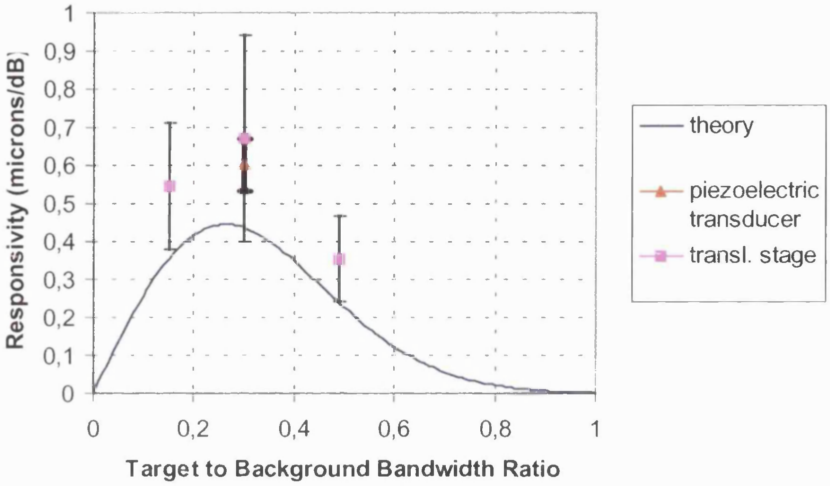


Figure 8.22 - Comparison between bandwidth optimisation theory and experiments

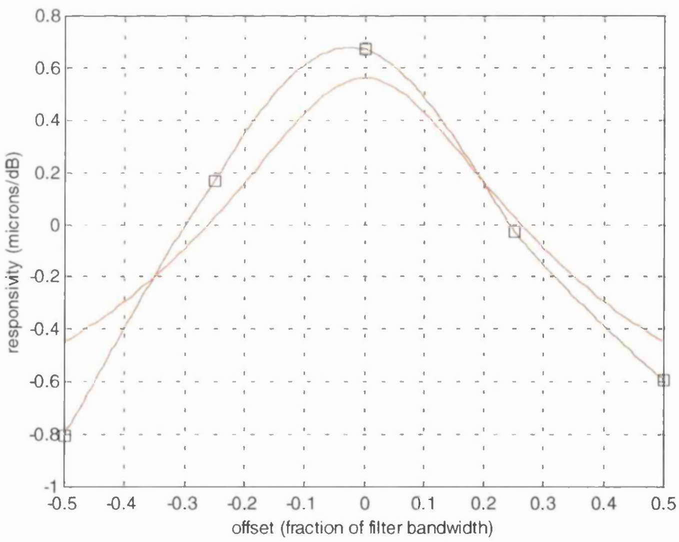


Figure 8.23 - Comparison between central wavelength offset theory and experiments (with squares)

Interestingly, we can observe in fig. 8.23 that a good positive responsivity can be obtained in a range of offsets that corresponds to approximately one third of the filter bandwidth. By thinking of a single line target scanning the filter spectrum as another broad target with a bandwidth equal to the scan range, this value of 1/3 agrees with the findings from the bandwidth ratio theory and experiments presented in this chapter.

8.4. Conclusion

This chapter presented the modelling and performance prediction of the IPSS approach. A simplified model was derived analytically and graphically, which was valid for spectrally narrow targets, and under additional assumptions. This model provided a good insight into the mechanisms affecting the performance, and its main conclusions were: for a given filter, the responsivity increases as the target bandwidth decreases; given a target, there is an optimum target to filter bandwidth ratio of approximately 0.3, which yields maximum responsivity. As an analytical derivation including the effects of detailed characteristics of components on the performance was not possible, a simulation model was developed using the understanding of the behaviour of the detection system, given by the simplified model. This extended model was capable of predicting the behaviour of the experiment with both coherent and partially coherent light sources. The uncertainty in power measurements and the variable current used to vary the SCR in the LED and RCLED experiments were the main causes of the differences found between phase step shift theory and experiments. An important finding taken from the use of the simulation model is that, for narrow targets, the target spectral shape does not influence the results. On the other hand, when considering spectrally larger targets, the ones modelled as a Gaussian spectrum produced higher responsivities than the Lorentzian targets. The simulation model also accounted for features such as the polarisation of the He-Ne laser and the non-flatness of the blackbody curve of the tungsten halogen lamp, and allowed the confirmation of effects such as the observability cut-off that occurs at high SCRs. The simplified model was also used to validate the experimental results from chapter five, concerning the optimum target to filter bandwidth ratio, and the effects of the target to filter central wavelength offset, both producing a very good agreement between experiment and theory. Now, the models developed have been validated, and their use is recommended for future design of detection systems employing the IPSS technique and algorithm.

The simulation model developed for performance prediction purposes can be used to find an optimal interference filter spectral shape that would optimise the responsivity. It is envisaged that this shape will be heavily dependent on the target spectral shape. This assessment is left as a suggestion for future work.

Chapter 9

Conclusions

This final chapter is a summary of the most important findings of the research work described in this thesis. It starts with a brief summary of conclusions from previous chapters, included here as a reminder to facilitate the reading of the rest of the chapter. The next section, 9.2, states the achievements and limitations of the research work. After reviewing the thesis, the chapter discusses the IPSS technique, presenting its advantages and limitations in section 9.3. The last section describes possible directions for future work.

9.1. Discussion of chapter conclusions

9.1.1. Chapter one – Introduction

This first chapter showed, in the motivation section, that coherence-based techniques are well suited to applications where dim, coherent targets must be detected in brighter, incoherent backgrounds. A list of applications of such techniques is given, such as defence, remote sensing, biological imaging, forensic science, satellite communications and astronomy. The chapter also presented the aims of the research.

9.1.2. Chapter two – Review of relevant detection techniques

In chapter two, a review of research work considered relevant for comparison with IPSS was presented. It was shown that spatial filtering depends heavily on the detection scenario, being of limited usefulness on its own, and temporal (or velocity) filtering is impaired by image registration misalignments. Track-before-detect (TBD) currently represents the state of the art for systems based solely on the intensity. The best sensitivity (minimum detectable signal-to-clutter ratio) for intensity-based techniques found in the researched literature was –6.6 dB, but the processing time is too long for real time applications (20 sec). There is a consensus amongst researchers that the evolution of signal processing techniques has reached a point where little improvement in sensitivity can be obtained with further intensity-based processing, and that only the introduction of other dimensions in this processing can achieve better results.

Dual band processing has been extensively used in missile warning systems, and has an improvement potential if the problems with inter-band background correlation are circumvented. Multispectral techniques, in their turn, can achieve better sensitivities, but imply the use of complex algorithms to remove the effects of different temperatures of the objects in a scene. Another inconvenience of these techniques, shared by hyperspectral approaches like Fourier transform spectroscopy (FTS), is the very large amount of data they generate, elevating the processing costs and, often, increasing the time required to produce the information required. In systems using spectral information, there is a trade-off between the use of *a priori* knowledge, and flexibility. A system employing *a priori* knowledge about the expected targets can yield better sensitivities with reduced processing effort, while those that do not use this knowledge display worse sensitivities, but have broadband operation and, hence, are more flexible. On the other hand, narrowband systems are less susceptible to interfering signals or features.

FTS systems present, under specific constraints, the well-known Fellgett and Jacquinot advantages, when compared to dispersive spectroscopic systems. Additionally, FTS has the advantage of wide bandwidth. Its use has been combined with background subtraction techniques with relative success when applied to the remote sensing of gases. Alternatively, there is a set of coherence-based techniques, which do not perform the Fourier transformation of FTS, and can sense a change in spectrum without directly measuring it. As an example, Optical Transform Image Modulation (OTIM) techniques [Fre72] have shown high sensitivities using pre-detector techniques. The work reported, however, did not include intensive post-detector signal processing.

The Coherence Measurement technique [Dru90] reviewed in chapter two introduced a modification to FTS, where path difference was varied just in the vicinity of a feature in the interferogram. A drawback of the approach comes from the need to measure the amplitude of the self-coherence function, which is impaired by uncertainties in path difference.

9.1.3. Chapter three – Theory

The third section of this chapter describes the technique we have named Interferogram Phase Step Shift (IPSS). The technique as a whole was invented by French [Fre93], and we have developed a signal processing algorithm to implement it [Cou99]. It is shown that IPSS utilises the same short scan as Coherence Measurement, both generating less data and being faster than FTS. Unlike Coherence Measurement, which measures the amplitude (real part) of the interferogram at a specific path difference, IPSS measures the displacement, in the path difference axis, of a phase feature that occurs in the interferogram. Due to this, IPSS does not require the calibration tasks needed by that method. This major difference makes IPSS, unlike the other techniques, immune to detector non-linearities, as well as to any other distortions affecting both target and background interferograms in the same way. One disadvantage of IPSS is, however, its narrowband operation, requiring *a priori* knowledge about the target to be detected.

9.1.4. Chapter four – Coherence length measurements

The measurement of coherence length of some of the light sources used in the experiments in chapters five and six gave a good insight into the experimental artefacts affecting the proper recording of an interferogram using a direct method in a Michelson interferometer. Most of these artefacts only affect long interferograms, which were used in the calculation of coherence lengths. The only artefact affecting the short IPSS interferograms was the additional modulation in the fringe signal arising from the stepwise speed of the translation stage, which had to be removed by software filtering. One important finding was that Wolf's formula for coherence length (eq. 3.6) showed higher immunity to noise, when compared to Mandel's (eq. 3.7). The measurement of coherence length of RCLEDs, apart from demonstrating that these devices have a coherence length that depend on the numerical aperture of the collection optics, showed that their coherence lengths fill an important gap in the coherence length spectrum, in between those of standard LEDs and of semiconductor lasers.

9.1.5. Chapter five – Non-imaging detection experiments

The linearity characteristic of the path difference driving schemes used in this chapter showed that only the results obtained with the translation stage were adequate for comparison with results predicted using the developed theory. On the other hand, these were the results that presented the highest level of noise, due to the stepper motor that was used to drive the stage. The fast piezoelectric transducer was the arrangement with the lowest noise level, as it produced high carrier frequencies (approximately 48 kHz), away from detector noise. Due to the transducer lack of linearity, the method displayed a relatively low responsivity. As the sensitivity (defined in chapter one) depends on the responsivity (defined in chapter eight) and on the noise level, this arrangement was not the one with the best sensitivity. Additionally, the update rates of the piezoelectric transducer of ~ 500 Hz represents the best alternative for real time applications, among the schemes tried.

The results obtained with the three different arrangements (translation stage, and two piezoelectric transducers) were in reasonable agreement, the differences being due to the non-linearities already mentioned. Very high sensitivities (-46.42 dB) were measured using a He-Ne laser as a target, while high sensitivities (from –13 to -30 dB) were demonstrated even for partially coherent targets. Light from a monochromator was detected at –30 dB, using a simplified detection criterion, consisting of finding the smallest SCR for which the phase step shift error bars were above zero. The sensitivity was found to increase with the coherence length of the light source, as expected in a coherence-based approach. Furthermore, it was found that there is an optimum target-to-filter bandwidth ratio, which maximises the responsivity, and this is around 0.3. The effect of an eventual offset between target and filter central wavelengths was measured as a beating effect seen in the interferogram, which can reduce the responsivity, or even make it negative for large offsets. When compared to an intensity-based method, in the laboratory, the IPSS sensitivity was 11 dB better.

9.1.6. Chapter six – Imaging detection experiments

Two different techniques were exploited in this chapter. The first formed Fizeau (line) fringes by tilting one of the interferometer's mirrors, and performed detection of targets without forming images. The phase step that occurs in the fringe carrier and is used by IPSS was shown as an image of line fringes, recorded with a CCD camera, and appeared as a change in the fringe spatial frequency. The sensitivity obtained for a laser target was high (-36 dB), but 10 dB worse than the temporal method, due to problems with background subtraction and with laser speckle. With partially coherent targets, the performance was better than that of the temporal solution, with a sensitivity of –33 dB to detect radiation from a monochromator. Additionally, this arrangement had the advantages of not suffering from the path difference drift seen in the temporal technique, and of using no moving parts. The results obtained were in good agreement with those obtained with the temporal approach, presented in chapter five.

The second set of experiments was concerned with coherence imaging. Firstly, some experiments done by other workers in the past were replicated to supply sensitivity data not included in the original papers. A computer algorithm was implemented to perform the detection of a laser source using the technique used by Sutton, named modulation recognition, which gave a sensitivity of –4.9 dB, for a point target. The result claimed by Hickman, obtained by subtracting two frames (-20 dB) was not reproduced, in similar experimental conditions. IPSS was the approach with the best sensitivity, with an advantage of 17.4 dB with relation to the other techniques for a point target. The penalty of its use is the large number of frames required to form a coherence-based image. When detecting extended targets, some approaches (including IPSS) did not work, but those who did gave better sensitivities than for point targets.

9.1.7. Chapter seven – Algorithm development and comparison

The spectral analysis of the interferograms acquired in the experiments confirmed the finding from chapter five that the translation stage was the scheme with the highest level of noise. The presence of noise spectral components close to the frequency of the fringe carrier required the use of a sharp roll-off software filter, the choice being an elliptic filter, due to its sharpest roll-off, among the filters studied. Four algorithms were compared in terms of their Receiver Operating Characteristics (ROC), of which IPSS displayed the best performance in terms of probabilities of detection and false alarm, both for small signals and in the whole SCR dynamic range. A computer program was developed to speed up this comparison, loading all interferograms to be analysed at once, and plotting the ROC curves from the resulting data, calculated using Hypothesis Testing.

It was shown that a laser target with an SCR of –26 dB was detected using IPSS with a probability of detection of 90%, and a probability of false alarm of 2.10^{-4}. These figures, for example, make IPSS 14 dB more sensitive than the results published by Schwarz [Sch96], also characterised using ROC curves. Concerning the sensitivities stated as minimum detectable signals, IPSS had better sensitivities than most approaches reviewed, including Coherence Measurement. The exceptions are the report made by French [Fre88], which did not give details about the experiments, the work by Duffy [Duf89], which employed a subjective detection criterion and the work by Montgomery and others [Mon00], which required intensive computing resources.

The chapter also presented the limitations of the IPSS algorithm, consisting of the detection of false frequency spikes and, in some cases, no detection at all. The algorithm detected the correct spike in 92% of the cases studied, which was considered a good performance, given the limitations of the hardware available to record the interferograms.

9.1.8. Chapter eight – Modelling and performance prediction

In this chapter, a novel, advanced theory was presented, which predicted the experimental results with good accuracy. In agreement with experimental observations in chapter five, the theory showed that:

I. Given an optical filter, the narrower the target spectrum, the larger the responsivity;

II. Given a target to be detected, there is an optimum target-to-filter bandwidth ratio, which depends on the target spectrum, as well as on the path difference dynamic range, and equals approximately 0.3;

III. The shape of the shift vs. SCR curve, in the case of narrow targets, is given by the spectrum of the optical filter. If this spectrum is assumed to be a top-hat function, the curve will have an "inverse-sinc" shape;

IV. Concerning the spectrum of the target, if it is not considered a "narrow target", Gaussian targets produce larger responsivities than Lorentzian targets having the same FWHM;

V. An eventual offset between the target and optical filter central wavelengths produces a reduction in responsivity, which might even reach negative values for offsets larger than 25% of the filter bandwidth.

A simulation model was presented, which includes the effects of the polarisation states of the light sources on the fringe visibilities from target and background. The theoretical results were in good agreement with those from the experiments employing the translation stage, especially those using the He-Ne laser as a target. The differences are mainly due to uncertainty in SCR measurement (approximately $\pm$ 1.5 dB), and to the variable supply current used in the LED and RCLED experiments. Considering this latter uncertainty, the agreement can be considered to be very good.

9.2. Achievements and limitations of the work described in this thesis

This section describes the main achievements and limitations of the research performed to pursue the aims described in chapter one. The pros and cons of the IPSS technique itself are described in the next section. The list is itemised by subject, and each topic starts with statements about the achievements, followed by related limitations, if any.

I. A novel algorithm to recover the phase step in the interferogram was developed - The use of: an interference filter to create a marker in an interferogram, an interferometer to create this interferogram, a short path difference scan, unlike FTS, and a phase step as a marker in the interferogram, were invented by French [Fre93], being object of a patent belonging to the UK Ministry of Defence. Optical arrangements implementing some of these ideas have been presented by Sutton [Sut82] and Drum [Dru90]. This thesis includes, as beneficial additions to previous optical implementations of the technique, the development of a novel algorithm comprising the digitisation of the interferogram signal and the extraction of a feature in the interferogram using tools adapted from the fields of telecommunications and signal processing. Furthermore, it is believed that this is the first time that intense digital post-detector signal processing is used in conjunction with optical pre-detector processing in a coherence-based approach. The algorithm presented an error rate of approximately 8%, considered to be low, mainly because of experimental artefacts at the edges of the interferograms acquired;

II. Algorithm comparison - The research demonstrated that the IPSS algorithm had the best performance among four algorithms studied, which were thought to be among the best for the task of coherence detection. The comparison used results generated without any manual intervention. However, it did not include all the experimental data generated during the research, which would require a study based on neural networks to assess the different outputs (figures from the ROC curves) caused by different inputs (several groups of experimental data, each one acquired under specific conditions);

III. Very high detection sensitivities (–46 dB) were demonstrated – These were measured when detecting a coherent light source. High sensitivities (–30 dB) were also measured with partially coherent sources, such as LEDs, RCLEDs and light from a monochromator. It can be said that these sensitivities are high because of the comparison with the intensity-based method in chapter five, with other approaches which had experiments replicated in chapter six, and with figures from the literature, included in chapter seven. The results were affected by limitations in the optical and electronic components used in the experiments. Some examples were: problems in light collimation caused by the optics created a high SCR uncertainty; imperfections in the optics that prevented the use of IPSS to image spatially extended targets; errors in phase step shift caused by the lack of repeatability and linearity of the devices used to vary the path difference; 100 Hz noise caused by the use of tungsten halogen bulbs powered from the AC supply.

The sensitivities obtained were measured in the laboratory, not including the effects of the atmosphere, which are discussed in the next section. Furthermore, the investigation was limited to stationary targets;

IV. Agreement between the results of different experiments - The detection experiments comprised two different methods, one with temporal and another with spatial interferograms, and the agreement between the results of both methods was good. The phase step (or frequency spike) shift vs. SCR characteristic was demonstrated utilising four different path difference driving arrangements, with reasonable agreement between their results;

V. Phase step - The existence of a phase step in the interferogram, due to a characteristic of the optical filter, was demonstrated through temporal signals and through images showing it;

VI. Coherence imaging - To the best knowledge of the author, this thesis presents the first demonstration of imaging based on the coherence contrast between target and background, relying on a signal processing algorithm, and achieving high sensitivity. A limitation of the demonstration was the relatively high number of frames (180) required to obtain one update;

VII. A novel theory was developed for modelling and performance prediction – This theory successfully predicted the experimental results concerning the responsivity of the IPSS approach, as well as its variation with the target-to-filter bandwidth ratio and central wavelength offset. The theory was complemented by a simulation model that predicted, with reasonable accuracy, the amount of phase step shift in the detection of different light sources. The model included most physical mechanisms affecting the results, and was quite flexible, allowing for arbitrary optical filter spectra and for two different types of target spectra. Concerning limitations, no success was obtained in the derivation of the self-coherence function of a Lorentzian target, being filtered by the optical filter (appendix VI). Additionally, the theory and simulation model did not account for the effects of noise, which, for comparison purposes, were considered through the experimental error bars;

VIII. ROC curves – It is believed that this thesis includes the first characterisation of a coherence-based detection system using Hypothesis Testing and Receiver Operating Characteristic (ROC) curves. As a limitation, the calculation of the curves assumed a Gaussian distribution of the phase step shifts without a rigorous demonstration of its validity;

IX. Coherence length measurements – In chapter four, the variation of coherence length of RCLEDs with the numerical aperture of the collection optics, was investigated experimentally for the first time [Cou02], representing a valuable contribution to the development of applications employing these devices.

9.3. Advantages and limitations of the IPSS technique

Undoubtedly, the principal advantage of the IPSS technique is its very high sensitivity. When compared to intensity-based detection systems, this advantage is explained by the correspondence that exists between the SCR and a Student t variable, used to characterise the phase step shift, as calculated in chapter seven. This variable is equivalent to a Signal-to-Noise ratio in conventional systems, but, in IPSS, values larger than one (e.g. 2.6) correspond to very low SCRs (e.g. –30 dB). When compared to dispersive spectroscopic techniques, IPSS incorporates the advantages known to exist in FTS systems (Fellgett and Jacquinot). Compared to FTS, and as happens with Coherence Measurement, the technique performs shorter path difference scans, allowing higher update rates, SNR gain by integration, and generating much less data. Furthermore, the ability to sense a change in spectrum without performing a Fourier transform reduces processing time and effort.

If compared to Coherence Measurement, IPSS is concerned with the measurement of a temporal or spatial delay (the phase step shift), while Coherence Measurement, as its name indicates, measures the increase in the degree of coherence at the path difference where the interferogram has its first minimum. This measurement is affected by amplitude noise in the detector and electronics, detector non-linearities, plus any factors influencing the visibility of the interference fringes obtained. As the self-coherence functions of target and background radiation are up-converted in an interferometer, and then narrowband filtered and demodulated in phase (using a computer algorithm in this case), IPSS inherently removes these types of noise. Thus, as IPSS does not need to measure amplitudes, it is not impaired by these factors, representing an advantage that reduces the measurement errors and simplifies the signal processing.

One problem to be circumvented by detection systems exploring spectral signatures is the ability to extract the desired information from the large volume of data available for processing. Multispectral systems, for example, generate four-dimensional data (two spatial, one temporal and one spectral dimension), in many cases collecting much more data than it would be actually necessary. IPSS uses prior knowledge about the expected target to collect this data selectively, reducing the amount of data and optimising the processing chain. Not only the wavelength of the expected target is considered, but its spectral width (coherence length) is considered as well. Hence, the use of *a priori* knowledge makes IPSS more sensitive, faster, and simpler than other systems relying on spectral signatures.

Another important advantage of IPSS is the use it makes of the background radiation (clutter). In conventional systems, clutter represents the factor that degrades the performance most severely. Nevertheless, IPSS uses the background radiation through an optical interference filter to form a specially shaped interferogram from it, actually taking advantage of this clutter to use it as a natural illuminator. If the spectral density is maintained, intensity fluctuations within the filter band do not alter the reference position of the phase step in the interferogram, making the approach immune to temporal clutter.

A final advantage comes from another comparison with FTS. In FTS systems, an additional laser is required to supply a path difference reference, used by the Fourier transformation process. Furthermore, a white light source is used as a zero path difference (ZPD) reference, allowing the acquisition of single-sided interferograms and reducing acquisition times. IPSS requires neither of these resources. The white light reference is not required due to the repetitive fashion in which path difference scan operates in IPSS. Given an optical filter, the path difference region to be scanned is always the same, where the first minimum in its self-coherence function occurs. This fact also explains why a reference laser is not necessary. If the path difference driving scheme has good repeatability, ensuring that each scan starts at the same path difference, with negligible drift, the IPSS approach is only concerned with the *relative* position of the phase step, without worrying about the *absolute* path difference value where it occurs.

As for the limitations of the IPSS technique, possibly the most significant limitation comes from the use of *a priori* knowledge about the expected targets. As discussed previously in this chapter, the use of a narrowband filter at the entrance of the detection system reduces its flexibility to detect targets in different spectral bands simultaneously. Even if the correct filter is employed, eventual deviations in the central wavelength of the expected radiation may cause loss of sensitivity due to the target-to-filter offset measured in chapter five and studied in chapter eight. The theoretical model developed showed that offsets of ± 10% of the filter bandwidth reduce the sensitivity in 50%. This may become an advantage if this sign reversal is used to create an additional modulation, as described in chapter eight.

In an attempt to minimise the problem caused by the target-to-filter wavelength offset, one may think of increasing the filter bandwidth. However, by doing so, another problem may arise, which is the loss of sensitivity due to the presence of more than one coherent feature within this bandwidth. Although the response of IPSS to multiple absorption or emission peaks has not been studied theoretically, assuming linearity and using the superposition principle, this effect will correspond to the sum of the responses to the individual coherent features. These features may be absorption or emission peaks, with different offsets, with relation to the filter's central wavelength. The weight of a response is proportional to its offset with relation to the filter central wavelength. Therefore, an absorption and an emission feature, with the same offset, may cancel out, and this may happen even with features at different wavelengths, provided their offsets are equal in size, but with opposite signs. Two or more features of the same type (absorption or emission) will add their phase step shifts, according to the weights previously mentioned.

The limitation of being a narrowband system can be mitigated via the use of filter wheels and employ time multiplexing, or through arrays of filters at different bands, employing spatial multiplexing. These measures would give the detection system the flexibility to detect features at different wavebands.

Another limitation of the technique is its inability to detect ultra-short pulses, from either lasers or other light sources. The shortest detectable pulse width is given by the processing time required to declare detection. In a temporal approach with a scan rate of a few kHz, obtainable with currently available piezoelectric transducers, this time will be in the order of milliseconds. In a spatial approach, like the one used in chapter six, this time will be the time taken to locate the phase step, which depends on processing speed. Luckily enough, the current trend in electronic warfare is to evolve from pulsed to continuous wave (CW) lasers to reduce their detectability by laser warning systems, which operate using the laser rise time signature [Der99]. The detection of a train of pulses, depending on the pulse interval, may be viable by time averaging.

Finally, it is important to discuss the eventual degradation of IPSS performance due to atmospheric effects. The propagation of optical beams in the atmosphere is a complex subject, with several active research areas. Although a rigorous investigation of these effects is outside the scope of this thesis, an initial analysis of the problem is given here. The optical beam coming from the target will arrive at the IPSS receiver after propagating through the atmosphere or another scattering or turbid medium, depending on the application. In the atmosphere, three main effects on the radiation from the target are envisaged:

a) Beam attenuation – It is expected that attenuation based on absorption and scattering by molecules and aerosols in the atmosphere would equally affect target and background radiation, not altering the SCRs and, hence, the detection sensitivities. In the case of the hot CO_2 blue spike feature, presented as an example in chapter one, attenuation by atmospheric effects turns out to be an IPSS advantage, as "*... hot CO_2 emissions propagate with much less attenuation than in atmospheric bands. This effect is pronounced in the 4.3 μm region of CO_2 absorption.*" [Zis93]. Hence, it is expected that, at least in this waveband, target radiation will be less affected than the background, increasing the SCR;
b) Scintillation – Produced by amplitude fluctuations resulting from the interference between different portions of the beam that propagated through atmospheric eddies with different refractive indices. Optical turbulence is the mechanism behind scintillation, which causes

temporal and spatial phase de-correlation of the wave front, and is usually described by the power spectral density of the phase fluctuations, using the Kolmogorov model [Smi93]. The spectrum of turbulence has been modelled and measured by several workers using Shack-Hartmann wave-front sensors, for example by Nicholls, Boremann and Dainty [Nic95]. Measurements done by other workers showed that the spectrum of turbulence fluctuations has a time scale of 10 sec, and its components at 100 Hz have amplitudes 80 dB below the level observed at that time scale [Smi93]. If one considers this frequency as the turbulence bandwidth, a system providing path difference scanning frequencies larger than 100 Hz would allow the removal of the temporal effects of wave-front distortion by averaging;

c) Beam scattering – Possibly the most important atmospheric effect on IPSS comes from the loss of temporal coherence by Doppler spreading, which is an additional effect of scattering of the beam by atmospheric aerosols, apart from attenuation. As, currently, most systems employed to detect radiation do so by measuring its intensity, the work on atmospheric effects so far did not address the issue of loss of temporal coherence directly. Swanson and VanWinkle [Swa93a] performed an investigation of this loss via experiments with a laser beam crossing a tank filled with water and particles in suspension, at different concentrations. The work used a coherence length measurement method similar to FTS, employing a Fabry-Perot etalon. The laser had a wavelength of 514 nm and an optical bandwidth of 3 MHz, and its coherence length was calculated using Mandel's formula shown in chapter three (eq. 3.7). The main finding of the work was that the coherence length decreased with the increase in optical density, defined as the path length vs. beam attenuation coefficient (effect described in a)). The coherence length in clear water was measured to be 22 m, decreasing to 10 m for an optical density of one, and to 4.2 m if this density is 2.6.

The experiments were not supported by any theoretical studies, hence their results cannot be directly extrapolated to other situations in which different wavelengths, optical bandwidths and particles in suspension are found. However, some conclusions may be obtained by calculating the responsivity change of the IPSS approach in respect to target coherence length, using eq. 8.13, and considering the detection of the same laser target used in the experiments, for propagation in the same medium. Two cases may be considered:

I. The filter available for the detection system has a fractional bandwidth of 1%, which is an average figure in the spectroscopy industry. In this case, even at the maximum optical density measured by Swanson and VanWinkle (2.6), no reduction in responsivity occurs. It is worth noticing that this optical density corresponds to an attenuation of 93% of the target power. In this case, the responsivity would be limited by the filter;

II. Considering the use of a hypothetical filter, which fulfils the optimality criterion derived in chapter eight, based on the target-to-filter bandwidth ratio, considerable degradation occurs if the target coherence length is reduced by propagation. For example, if this length is reduced by 50%, the responsivity falls to approximately 6% of its original value. Unlike the first case, here the responsivity would be limited by the target coherence length. Nevertheless, an interference filter with an optical bandwidth of approximately three times the bandwidth of the laser used in the experiments (3 MHz or 2.6×10^{-6} nm) cannot be manufactured with current technology. The state of the art filters used in the telecommunications field, specifically in Wavelength Division Multiplexing (WDM) reach now a bandwidth of 50 GHz [Chu01].

To conclude the atmospheric propagation issue, the understanding of the effects of propagation on the coherence of radiation have been characterised neither theoretically nor experimentally. Although the real susceptibility of the IPSS approach to these effects can only be properly understood after such characterisation is performed, it is expected that even with degradation from the propagation medium, the IPSS approach would still display better sensitivities than most existing detection systems.

In conclusion to this section, from the advantages and limitations presented, it is clear that the IPSS approach is a powerful detection technique, being extremely sensitive and relatively simple to implement. As mentioned in chapter one, the work described in this thesis can be considered as fundamental research, as it was not aimed at any specific application. The thesis has demonstrated that further projects of applied research using the technique are worth pursuing. Some suggestions on how the work may be carried further are presented in the next section.

9.4. Work to be done

In the course of the research work, the progress towards the assigned aims could be compared with walking on a staircase, alternating periods of research to find different alternatives to the solution of a problem (a plateau) with its actual solution and consequent quick progress (a step up). Some of the alternatives where possibly worth pursuing, and some further steps may have been taken at the end of the staircase, but none of these where carried out due to lack of time and resources. This section briefly present some tasks to be performed in future research projects exploiting IPSS.

As mentioned in chapter three, the IPSS approach can be applied to any spectral band where components are available. One possible experiment would consist in its application to the middle IR band, to detect the CO_2 blue spike, described in chapter one. This would require IR coupling optics, an interference filter, an interferometer and radiation sources. The interferometer can be adapted from an existing FTIR spectrometer, with software changes. Background radiation could be obtained with a blackbody, and the target could be simulated with a gas cell containing room temperature CO_2, placed between another cell containing heated CO_2 and the detection system. Another possibility is the use of the 1.5 μm band, widely used in telecommunications. There are many types of semiconductor lasers to serve as targets, and WDM filters with narrow bandwidths and sharp roll off characteristics would give interferograms with easily detectable phase steps. One can even think of a coherence modulation communications system, where the signal would be coded as the phase step shift, caused by the variable power of a laser superimposed on broadband radiation. If free space optics is used, the latter could even be made from natural illumination. IPSS could also be applied to the development of radar or sonar receivers, and the skilled designer in these bands would have to devise how to implement its two main hardware components, the narrowband filter and the interferometer.

A system combining coherence and polarisation discrimination can be considered to have a polarisation-sensitive beamsplitter in the interferometer. As shown in chapter eight, this measure would give a polarised target an additional gain with relation to the unpolarised background, increasing the effective SCR rendered to the coherence-based system.

The simulation program developed in chapter eight to model the shift vs. SCR characteristic for an arbitrary optical filter shape can be used to optimise the filter's response such that the responsivity of the resulting system is maximised. The optimum shape would certainly depend on the spectrum of the expected target, and this dependence could be turned into a disadvantage if an unexpected target appears instead.

The computer algorithm used to measure the position of the phase steps in path difference can be improved in several ways. The phase step (and frequency spike) represents abrupt changes in frequency, happening at variable instants in time. Wavelet transforms are applicable in such situations, by producing a time-resolved spectrogram (similar to the one in fig. 7.1), which correlates spectral content and time. A neural network can be constructed to find the best wavelets to be used in different experiments, and the different groups of experimental data obtained can be used to train the network. In another possibility, the combination of instantaneous frequency and amplitude explored in chapter seven could be perfected by developing better weighting functions from the amplitude, to give a method uniting the accuracy of the frequency with the larger dynamic range of the amplitude. The complex correlation algorithm presented in chapter seven can be improved by developing better templates through the use of a large number of test interferograms and compensating for the phase deviations between them, caused by drift in the mirror moving hardware. Yet another idea applicable to locate the phase step, invented by French, would be passing the interferogram signal through a circuit that adds it to an amplified, delayed and inverted replica of itself [Fre93]. This circuit would provide an event marker, which is independent of the slope and spectrum of the original signal.

In the technique studied in this thesis, the interferogram was an electrical signal, obtained by adding the received radiation to a delayed version of itself, in an optical interferometer. This provided modulation of the received radiation in the form of fringes observed when path difference was varied in time (or spatially), and the sought after feature was located by applying a phase demodulation algorithm to a digitised version of the signal. As mentioned previously, the modulation followed by a demodulation can serve to remove or mitigate detector noise. Different modulation schemes, however, may be employed. French [Fre93] proposed the use of the interferogram signal, still with its DC level arising from the non-oscillatory terms in the General Interference Law (eq. 3.5), to modulate an FM carrier (at radio or microwave frequencies), such that, even in the minimum visibility region, a constant frequency deviation would occur. The processing required to find the location of the phase step would then be performed in this high frequency carrier, away from noise, and this would be demodulated afterwards. A microwave version of the IPSS processing would consist of using a coherent local oscillator (a laser) at an optical frequency capable of down-converting the optical signal to a convenient microwave intermediate frequency, where the phase step could be recovered by a phase-sensitive detector. Another modulation scheme was mentioned in chapter seven, where the reversal of the sign of the responsivity with the variation in target to filter central wavelength offset may be used to create a modulation in the position of the phase step, which could be used, for example, in a communications system.

In the approach studied in chapter six, using a CCD camera to measure fringe profiles, a study needs to be carried out concerning the number of lines to be averaged to form an IPSS interferogram. The use of low aberration optics, and higher quality interferometer components would yield a sequence of lines with reduced longitudinal relative displacement, allowing a larger number of lines to be averaged, and higher interferogram SNRs to be obtained.

Another important issue to be considered in the future is the application of the imaging mode of IPSS to moving targets. High frame rate cameras are required, such that the number of frames required to locate the phase step in the interferogram of the considered pixel (180 in chapter six) can be scanned within the update time required by the application. As an example, if 180 frames are required for one update, and 25 updates per second are needed, the minimum frame rate would then be 4500 per second. If this rate is achievable, than standard frame registration and tracking algorithms can be employed.

Finally, an IPSS system operating on a real-time basis may be designed and built. Such a system would present the following main differences with relation to the laboratory arrangements used throughout the thesis:

I. Path difference driving scheme – A fast piezoelectric transducer, with position feedback;
II. Interferometer - The interferometer must be compact and rugged, to reduce its susceptibility to the environment. Current technology to do so include fibre, as well as monolithic interferometers, the latter consisting of a conventional interferometer with its optical components assembled in a monolithic unit [Ble99];
III. Narrowband filter - Some improvements to this filter may be implemented to give the system more flexibility, such as to employ liquid crystal technology to build a tuneable filter where a Fabry-Perot cavity is filled with a birefringent material, which can vary its refraction index with an external voltage. This would then allow the central wavelength to be varied, giving more flexibility to the approach to cope with the presence of targets with spectral lines at different wavelengths;
IV. Phase demodulation scheme – Designed to operate with a continuous signal, consisting of a sequence of successive scans, each one containing an interferogram, and the (undesired) effects of the transducer direction reversal, as measured in chapter five (see fig.5.3). The signal is sent to a PLL (Phase Lock Loop) implemented with a DSP chip, where its instantaneous phase and frequency are calculated. The PLL must have an appropriate bandwidth so that variations in signal frequency can be tracked. This bandwidth would also act as the software narrowband filter employed in this thesis;

V. Phase step shift measurement – Unlike the off-line system used in the experiments, keeping track of the phase step shift would involve registering at least four events: start of scan, reference path difference (position of the phase step with no target, acquired by pointing the sensor to a target-free region), current path difference, and end of scan. Detection is declared by observing the distance between the second and third events. A pulse width (PWM) or pulse position (PPM) system may be used to encode the temporal variations of the phase step shift. Averaging successive scans may also be helpful to increase the SNR, observing the caveats described in chapter five;

VI. Processor - A microcontroller or microprocessor would perform the tasks of controlling the other elements in the system, such as the mirror scan speed, man machine interface, etc.

The author believes that the work described in this thesis has obtained relevant progress in the body of knowledge of coherence-based techniques, that the work has quantified and optimised the performance of the IPSS system, and that the IPSS technique, together with the IPSS algorithm developed as part of the work, deserves recognition and further application-oriented research.

APPENDIX 1 – PROGRAM LISTINGS

(All programs written in Matlab version 5.3 language)

1. Program used in chapter 4 to calculate coherence lengths

```
% THIS PROGRAM CALCULATES COHERENCE LENGTHS
% load interferogram (1024 samples)
clear
cd C:\UCL\expdata\.pcnfs\monoc
load pos1.txt;
a3=pos1;

% removing DC level
a2=a3-mean(a3);

% rectifying
a1=abs(a2);

% changing horizontal axis
i=1:1024;

% selecting useful region
start=200;
finish=700;
a=a1(start:finish);
[Y2,I2]=max(a);

% lowpass filtering
[c,b]=butter(5,100/512);
d=filter(c,b,a);

% adjust interferogram to have the same number of samples to either side of the maximum
[Y1,I1]=max(d);
centre=(finish-start)/2;
if I1>=centre,
    e1=(finish-start)-I1;
end
if I1<centre,
    e1=I1;
end
e=d(I1-e1+1:I1+e1+1);
[Y,I]=max(e);

% calculating coherence length using Wolf formula (eq. 3.6)
j=[-e1:e1];
f=sqrt(sum((j.*e).^2)/sum(e.^2));

% converting coherence length to microns
% cv=conversion factor
cv=2*0.0977;
disp('Wolf='),disp(cv*f)
plot(cv*j,e*Y2/Y1,'b-',cv*j,a2(I2-e1+start:I2+e1+start),'r-');
grid on;
% CL using Mandel formula (eq. 3.7)
f1=Y^(-2)*sum(e.^2);
disp('Mandel='),disp(cv*f1)
```

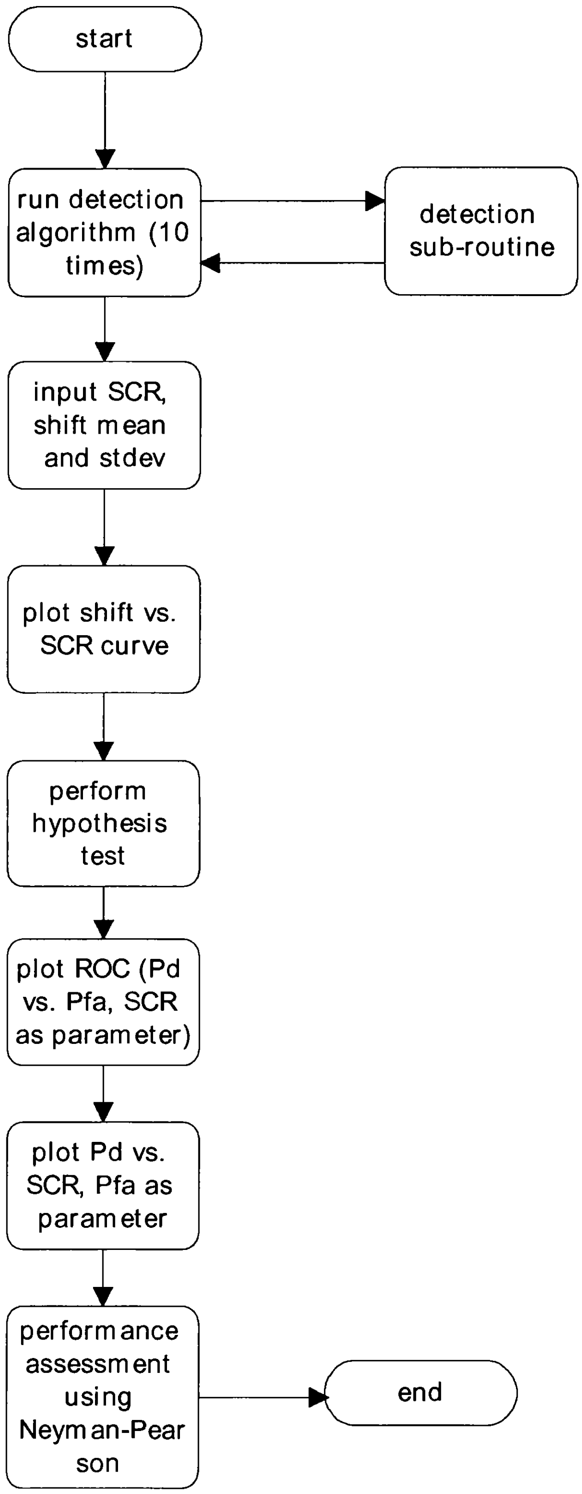
2. Program used in chapter 4 to perform interferogram spectral estimation

```
% This program obtains the PSD from an interferogram
clear
% load interferogram
cd c:\uclcds\.pcnfs\monoc
load pos1.txt;

% selecting a region without fringes to estimate the noise
Y2=pos1(600:700);
% removing noise from interferogram (this step is required only if the SNR is too low)
%[b,a]=butter(4,[0.99*0.578 1.01*0.578]);
%Y3=filter(b,a,Y2);

% applying a hamming window
Y3=Y2.*(hamming(101))';
% calculate interf. PSD
Y=fft(Y3,1024);
Pyy=Y.*conj(Y)/1024;

% removing DC component of PSD, which causes errors when calculating the frequency axis
A=1:20;
Pyy(A)=0;
% generate baseband frequency axis
% FFT length=1024/ sampling freq=1024
f=(0:511)/1024*1024;

% converting baseband Hz to optical microns^-1
% f2=f/(2*mirror speed in microns/sec)
f2=f/(2*100);
% converting wavenumbers in microns^-1 to wavelengths in nanometres
f3=1000./f2;
% normalising the spectrum
Pyy=Pyy/max(Pyy);
% plot of PSD vs. wavelength in nanometres
plot(f3,Pyy(1:512))
axis([600 720 0 1.2])
```

3. Programs used in chapter seven to generate the ROC curves

3.1. Main program

3.1.1. Flowchart



3.1.2. Program listing

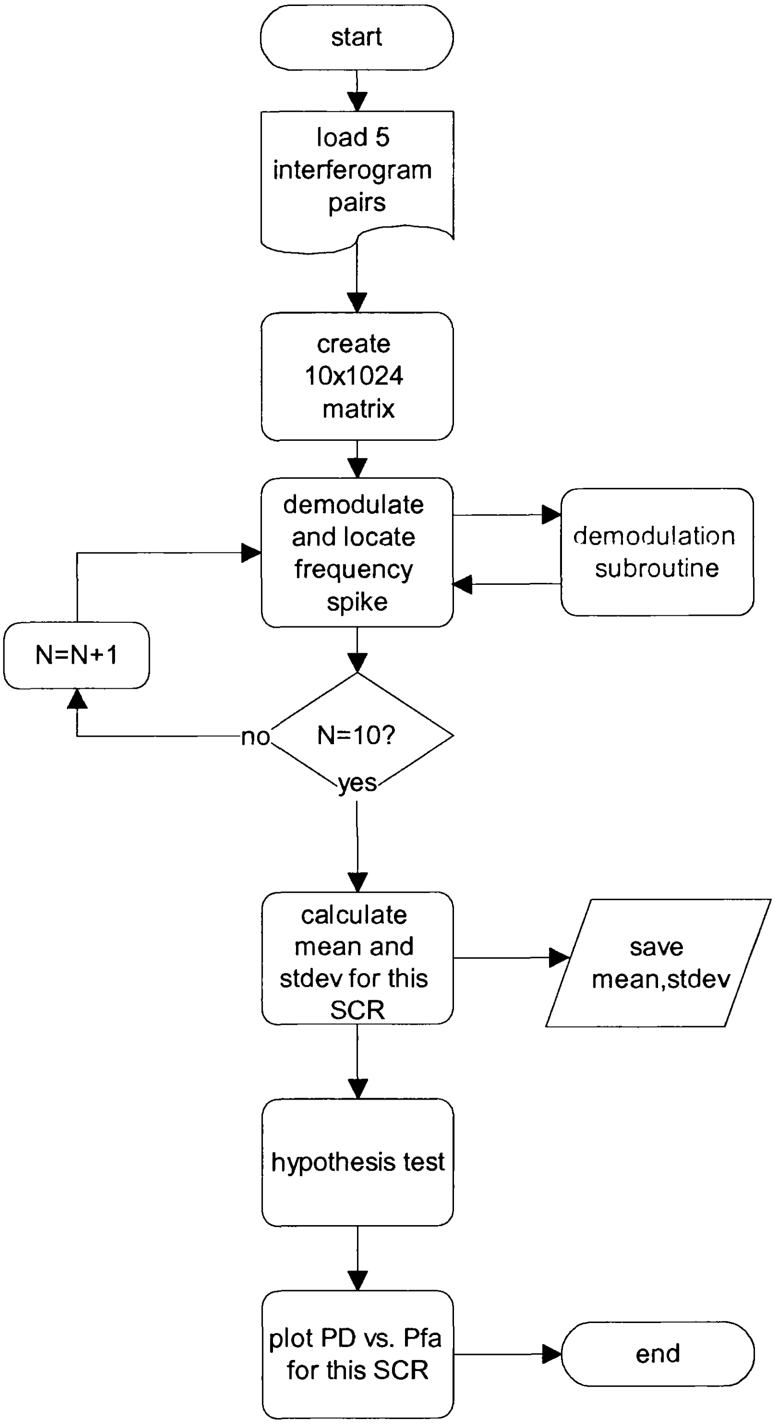
```
clear
cd c:\matlabr11\work
delete results.mat  % deleting results previously stored
% processing 5 interferogram pairs per SCR
detection11
detection12
detection13
```

```
detection14
detection15
detection16
detection17
detection18
detection19
detection20
% input scr vector
scr=[-50.03 -45.41 -40.64 -35.75 -30.38 -25.45 -20.38 -15.78 -12.43 -10.69];
total=[aveshift1 aveshift2 aveshift3 aveshift4 aveshift5 aveshift6 aveshift7 aveshift8 aveshift9
aveshift10];
stdev=[stdshift1 stdshift2 stdshift3 stdshift4 stdshift5 stdshift6 stdshift7 stdshift8 stdshift9
stdshift10];
%plot shift vs. SCR
% sample number to microns conv. Factor= 0.0293; wavelength=0.6328
errorbar(scr,0.0293*total/0.6328,0.0293*stdev/0.6328);
xlabel('Signal-to-Clutter Ratio (dB)');
ylabel('Marker Displacement (wavelengths)');
grid on;

% hypothesis test
% H0: shift=0, target not detected
% H1: shift>0, target detected
% calculation of test statistics (sample size=5)
t41=aveshift1./stdshift1./sqrt(5);
t42=aveshift2./stdshift2./sqrt(5);
t43=aveshift3./stdshift3./sqrt(5);
t44=aveshift4./stdshift4./sqrt(5);
t45=aveshift5./stdshift5./sqrt(5);
t46=aveshift6./stdshift6./sqrt(5);
t47=aveshift7./stdshift7./sqrt(5);
t48=aveshift8./stdshift8./sqrt(5);
t49=aveshift9./stdshift9./sqrt(5);
t410=aveshift10./stdshift10./sqrt(5);
% receiver operating characteristic
% Pd vs. Pfa
% varying threshold
thres=-100:0.1:100;
% calculating probability of detection in Student's t curve
Pd1=1-tcdf(thres-t41,4);
Pd2=1-tcdf(thres-t42,4);
Pd3=1-tcdf(thres-t43,4);
Pd4=1-tcdf(thres-t44,4);
Pd5=1-tcdf(thres-t45,4);
Pd6=1-tcdf(thres-t46,4);
Pd7=1-tcdf(thres-t47,4);
Pd8=1-tcdf(thres-t48,4);
Pd9=1-tcdf(thres-t49,4);
Pd10=1-tcdf(thres-t410,4);
Pfa=1-tcdf(thres,4);
% plot of Pd vs. Pfa for some selected SCRs
figure(2)
semilogx(Pfa,Pd1,'b-',Pfa,Pd7,'r-',Pfa,Pd6,'g-',Pfa,Pd10,'k-',Pfa,Pd8,'m-');
xlabel ('Probability of False Alarm');
ylabel ('Probability of Detection');
axis([1e-5 1 0 1]);
grid on;
```

```
% Pd vs. SCR for the Pfas of 0.1, 0.01 and 0.001
Pfa1=1e-1;
thres1=tinv(1-Pfa1,4);
Pd11=1-tcdf(thres1-t41,4);
Pd21=1-tcdf(thres1-t42,4);
Pd31=1-tcdf(thres1-t43,4);
Pd41=1-tcdf(thres1-t44,4);
Pd51=1-tcdf(thres1-t45,4);
Pd61=1-tcdf(thres1-t46,4);
Pd71=1-tcdf(thres1-t47,4);
Pd81=1-tcdf(thres1-t48,4);
Pd91=1-tcdf(thres1-t49,4);
Pd101=1-tcdf(thres1-t410,4);
Pdd1=[Pd11 Pd21 Pd31 Pd41 Pd51 Pd61 Pd71 Pd81 Pd91 Pd101];

Pfa2=1e-2;
thres2=tinv(1-Pfa2,4);
Pd12=1-tcdf(thres2-t41,4);
Pd22=1-tcdf(thres2-t42,4);
Pd32=1-tcdf(thres2-t43,4);
Pd42=1-tcdf(thres2-t44,4);
Pd52=1-tcdf(thres2-t45,4);
Pd62=1-tcdf(thres2-t46,4);
Pd72=1-tcdf(thres2-t47,4);
Pd82=1-tcdf(thres2-t48,4);
Pd92=1-tcdf(thres2-t49,4);
Pd102=1-tcdf(thres2-t410,4);
Pdd2=[Pd12 Pd22 Pd32 Pd42 Pd52 Pd62 Pd72 Pd82 Pd92 Pd102];

Pfa3=1e-3;
thres3=tinv(1-Pfa3,4);
Pd13=1-tcdf(thres3-t41,4);
Pd23=1-tcdf(thres3-t42,4);
Pd33=1-tcdf(thres3-t43,4);
Pd43=1-tcdf(thres3-t44,4);
Pd53=1-tcdf(thres3-t45,4);
Pd63=1-tcdf(thres3-t46,4);
Pd73=1-tcdf(thres3-t47,4);
Pd83=1-tcdf(thres3-t48,4);
Pd93=1-tcdf(thres3-t49,4);
Pd103=1-tcdf(thres3-t410,4);
Pdd3=[Pd13 Pd23 Pd33 Pd43 Pd53 Pd63 Pd73 Pd83 Pd93 Pd103];

% plot Pd vs. SCR
figure(3)
plot(scr,Pdd1,'b-',scr,Pdd2,'r-',scr,Pdd3,'k-');
xlabel('Signal-to-Clutter Ratio (dB)');
ylabel('Probability of Detection');
grid on

% performance assessment using the Neyman-Pearson criterion
% acceptable Pfa=10-2
disp(mean(Pdd2))
```

3.2. Sub-routine to calculate P_D and P_{FA} for each SCR (detection11.m)

3.2.1. Flowchart



3.2.2. Program listing

```
% data input (five pairs of interferograms)
cd C:\UCL\expdata\.pcnfs\controller
clear;
load lasmin84.txt;
load lasmin85.txt;
load lasmin86.txt;
load lasmin87.txt;
load lasmin88.txt;
load lasmin89.txt;
load lasmin90.txt;
```

```
load lasmin91.txt;
load lasmin92.txt;
load lasmin93.txt;
% forming a 1024 x 10 matrix
data=cat(1,lasmin84,lasmin85,lasmin86,lasmin87,lasmin88,lasmin89,lasmin90,lasmin91,lasmin
92,lasmin93);

% processing of the individual interferograms
for i=1:10;
  A=data(i,:);
  demod8  % detection algorithm
  % data output
  B(i)=I;
end

% calculation of shift statistics
nt=[B(1) B(3) B(5) B(7) B(9)];
wt=[B(2) B(4) B(6) B(8) B(10)];
shift1=[B(2)-B(1) B(4)-B(3) B(6)-B(5) B(8)-B(7) B(10)-B(9)];
aveshift1=mean(shift1);
stdshift1=std(shift1);

% saving average shift (aveshift) and standard deviation (stdshift) in file results.mat
save ('c:\MATLABR11\work\results','aveshift1','stdshift1')
```

3.3. Sub-routines to perform demodulation and shift measurement

3.3.1. Using IPSS (demod8.m)

```
% selection of useful region (to remove undesirable features at the edges of the interferogram)
u1=1;
u2=750;
c=A(u1:u2);
% extracting DC level
d=c-mean(c);

% bandpass filtering
% passband 440-500 Hz; stopband 0-400 and 540-5120 Hz; sampling frequency 10240 Hz;
% passband ripple 1 dB (max); stopband attenuation 50 dB (min)
[n,Wn]=ellipord([440 500] /5120,[400 540]/5120,1,50);
[b,a]=ellip(n,1,50,Wn);
g=filtfilt(b,a,d);

% frequency demodulation
% calculated carrier frequency 470 Hz
l=demod(g,470,10240,'fm');
% Removing edge spikes
for k=1:30;
l(k)=l(31);
l(u2-u1+1-(30-k))=l(u2-u1+1-30);
end
% position of the frequency spike
[Y,I]=max(abs(l));
```

3.3.2. Using minimum amplitude

```
% selection of useful region (to remove undesirable features at the edges of the interferogram)
u1=1;
```

```
u2=750;
c=A(u1:u2);
% extracting DC level
d=abs(c-mean(c));
% filtering high-frequency noise
[n,Wn]=ellipord(90/5120,150/5120,1,50);
[b,a]=ellip(n,1,50,Wn);
g=filtfilt(b,a,d);
% calculating the analytic signal
anasig1=hilbert(g);
% complex envelope
t=0:1/10240:(u2-u1+1)/10240;
ce1=anasig1.*exp(-j*2*pi*470*t);
% position of the minimum
[P,I]=min(ce1);
```

3.3.3. Using IPSS and minimum amplitude

```
% selection of useful region (to remove undesirable features at the edges of the interferogram)
u1=1;
u2=750;
c=A(u1:u2);
% extracting DC level
d=c-mean(c);
% bandpass filtering
[n,Wn]=ellipord([440 500]/5120,[400 540]/5120,1,50);
[b,a]=ellip(n,1,50,Wn);
g=filtfilt(b,a,d);
% frequency demodulation
l=demod(g,470,10240,'fm');

% amplitude of complex envelope
% calculating the analytic signal
anasig1=hilbert(d);
% complex envelope
t=0:1/10240:(u2-u1+1)/10240;
ce1=anasig1.*exp(-j*2*pi*470*t);
ce2=abs(ce1);
% combining frequency and amplitude
l2=abs(l).*(max(ce2)-ce2);
% Removing edge spikes
    for k=1:30;
    l2(k)=l2(31);
    l2(u2-u1+1-(30-k))=l2(u2-u1+1-30);
  end
% position of the frequency spike
[Y,I]=max(abs(l2));
```

3.3.4. Using complex correlation

```
% selection of useful Region (to remove undesirable features at the edges of the interferogram)
u1=1;
u2=750;
c1=A(u1:u2);
c2=C(u1:u2);
% extracting DC level
d1=c1-mean(c1);
d2=c2-mean(c2);
% bandpass filtering
```

```
[n,Wn]=ellipord([440 500]/5120,[400 540]/5120,1,50);
[b,a]=ellip(n,1,50,Wn);
g1=filtfilt(b,a,d1);
g2=filtfilt(b,a,d2);
% calculating the analytic signal
anasig1=hilbert(d1);
anasig2=hilbert(d2);
% complex envelope
t=0:1/10240:(u2-u1+1)/10240;
ce1=anasig1.*exp(-j*2*pi*470*t);
ce2=anasig2.*exp(-j*2*pi*470*t);
% locating region of minimum amplitude
[Y,J]=min(ce1);
% creating window centred at the region of minimum amplitude
H=hanning((u2-u1+1)/5);
x=1:length(H);
H1(x-(u2-u1+1)/10+J)=H(x);
H1(u2-u1+1)=0;

% applying window to no-target complex envelope
ce11=ce1.*H1;
noedge1=ce11(u1:u2);
noedge2=ce2(u1:u2);
% generating the cross-correlation function
cor=xcorr(noedge1,noedge2);
% locating the correlation peak
[Y1,I1]=max(cor);
% converting correlation lag to feature displacement
I=I1-(u4-u3+1);
```

4. Simulation model described in section 8.2

```
% Program to simulate the phase step shift
% in the target + background self-coherence function
% using an arbitrary filter spectrum and a Gaussian or Lorentzian target
clear
L=632.8; % wavelength
% input target bandwidth
d1=input('enter target FWHM in nm');
% entering arbitrary filter response
fil63311

% entering Planck curve
% M is the spectral radiance excitance in W/cm2.m
% light bulb blackbody temperature=3220K
M=2*pi*6.626e-34*(3e8)^2./(l./1e9).^5./(exp(6.626e-34*3e8./(l.*1.381e-23*3220/1e9))-1);
% Converting M to W/cm2.nm
M1=1e-9*M;

% filtered background
fb=ii.*M1;
% Converting bandwidths in nm to microns^-1
% (assuming symmetric spectrum)
d=1000*d1/(L^2);
% converting wavelengths in nm to wavenumbers in microns-1
k=1000./l;
L1=1000/L;
```

```
% normalised filtered background (in wavenumbers)
nfb=fb/trapz(-k,fb);

%normalised filtered target response
% Gaussian target (eq. 8.1)
gt=(0.9394/d).*exp(-2.7726*((k-L1)/d).^2);
nft1=(ii.*gt)/trapz(-k,ii.*gt);
% Lorentzian target (eq. A5.1)
lt=(2/pi/d)./(1+(2/d*(k-L1)).^2);
nft2=(ii.*lt)/trapz(-k,ii.*lt);

% generate path difference axis in microns, to plot interferograms
% FFT length=262144/ sampling freq=(632.8)^2 /(0.01*1000)[um]
f=(0:131071)/262144*L^2/10;

warning off

% Calculating Self Coherence Function for -50<SCR<0 dB
for s=-55:-5,
  %s=50:-0.1:0.1,
    PR=10.^(s/10);
  S1=nfb+PR*nft1;
  S2=nfb+PR*nft2;

  % FFT
  G1=ifft(S1,262144);
  G2=ifft(S2,262144);

  % selection of region of interest in the self-coherence function
  G11=G1(1:60*262144*10/L^2);
  G22=G2(1:60*262144*10/L^2);

  % determine phase step position
  Z1=abs(diff(unwrap(angle(G11))));
  [Y1,I1]=max(Z1(1:350));
  Z2=abs(diff(unwrap(angle(G22))));
  [Y2,I2]=max(Z2(1:350));
  A1(s+56)=I1;
  A2(s+56)=I2;
end

% plots – polarisation gain = 1.309
s1=-55:-5;
plot(s1,1.309*(A1-A1(1))/(262144*10/L^2),'r-',s1,1.309*(A2-A2(1))/(262144*10/L^2),'b-')
hold on
% experimental results
scr1=[-50.03 -45.41 -40.64 -35.75 -30.38 -25.45 -20.38 -15.78 -12.43 -10.69];
shi1=[0.1482 -0.202 -0.026 -0.107 0.536 0.624 0.718 1.5834 4.758 9.16];
err1=[0.087 1.305 0.08 0.422 0.482 0.638 0.68 0.98 0.666 0.63];
errorbar(scr1,1.5023*shi1,1.5023*err1)
hold off
```

5. Program to calculate the change in responsivity with target to filter central wavelength offset

```
% Program to assess theoretical variation in sensitivity
% due to target-to-filter detuning. Rectangular Background
% Gaussian or Lorentzian target with CWL offset
clear
% input bandwidths
L=674.8; % central wavelength
d1=input('enter target FWHM in nm');
B1=input('enter filter bandwidth in nm');
% Converting bandwidths in nm to microns^-1
% (assuming symmetric spectra)
d=1000*d1/(L^2);
B=1000*B1/(L^2);
% Path Difference given in microns
t=0.01:0.01:1.5/B;

% creating reference for shift
F1=(sin(pi*t*B))./(pi*t);
[Y0,I0]=max(diff(angle(F1)));

% Varying the offset and calculating self-coherence function for a %fixed PR
for O=-50:50;
  o=O/100*B;
  s=-8;
  PR=10.^(s/10);
  %Gaussian target (eq. 8.14a)
  G1=((sin(pi*t*B))./(pi*t*B))+PR.*exp(-3.559707*d^2*t.^2).*exp(j*2*pi*t*o);
  % Lorentzian target (eq. 8.14b)
 G2=((sin(pi*t*B))./(pi*t*B))+ PR.*exp(-pi*d*abs(t)).*exp(j*2*pi*t*o);
  % find minimum
  [Y1,I1]=max(abs(diff(angle(G1))));
  [Y2,I2]=max(abs(diff(angle(G2))));
  A1(O+51)=I1;
  A2(O+51)=I2;
end
O1=-0.5:0.01:0.5;
A3=max((A1-I0)/100);
A4=max((A2-I0)/100);
plot(O1,(A1-I0)/100/5.496,'r-',O1,(A2-I0)/100/5.496,'b-')

% comparing with the experiment
% input experimental data
ofst=[-0.5 -0.25 0 0.25 0.5];
resp=[-0.803058588 0.164879913 0.670707275 -0.028427948 -0.595567686];
resp1=interp(resp,50,2,0.5);
ofst1=interp(ofst,50,2,0.5);
hold on;
plot(ofst1,resp1)
hold off
```

APPENDIX 2 – LIST OF SIGNALS AND IMAGES USED IN THE RESEARCH

1. FILES (files marked as bold where used to generate results; all the others were used for calibration and adjustments)

File name	target	Filter CWL-FWHM (nm)	average number	amplifier			Sampling (kHz)	Drive method	Scan freq. Hz	Scan (μm)	Iris diam. (mm)	Exit lens	comments
				Gain dB	F_L (Hz)	F_H (Hz)							
Min1-8	He-Ne 632.8-0.002	633-11	variable	N	N	N	1	piezo	1,5 trian.	Variable	N	N	Setting-up
9-22	He-Ne	633-11	13	N	N	N	2,5	piezo	1,5	36,6	N	N	Drift assessment
23-30	He-Ne	633-11	variable	N	N	N	0,5	piezo	1,5	36,6	N	N	Shift calibration
31-33	He-Ne	633-11	variable	N	N	N	5	piezo	1,5	36,6	N	N	Drift
34-35	He-Ne	633-11	variable	N	N	N	5	piezo	1,5	36,6	N	N	Shifts
36-39	He-Ne	633-11	23 or 0	30-40	1	100	5	piezo	1,5	36,6	N	N	Amplifier test
40-64	He-Ne	633-11	variable	40	1	1000	5	piezo	1,5	36,6	N	N	Shift, 3 pairs/SCR
65-81	He-Ne	633-11	variable	40	1	1000	5	piezo	1,5	36,6	N	N	Shift, 2nd trial
82-185	He-Ne	633-11	variable	40	1	1000	5	piezo	1,5	36,6	N	N	5 pairs/SCR [Cou99], group 1
187-194	He-Ne	633-11	25-35	30	10	10000	5	transl. stage	150 μm/s	30	N	N	drift and repeatability
195-216	He-Ne	633-11	25-35	30	10	10000	5	t.stage	150	20	N	N	Shift, 1 pair/SCR
217-240	He-Ne	633-11	24	30	10	10000	2,5	t.stage	150	20	N	N	[Cou00], group 2
241-300	He-Ne	633-11	24	30	100	10000	2,5	t.stage	150	20	N	N	[Cou00], group 3
301-318	InGaAlP LED	651.9-36.2	1-3	30	100	10000	5	t.stage	50	20	N	N	Trial, filter on rotation stage
319-322	LED	651.9-36	1-30	30	100	10000	5	t.stage	50	20	N	N	Group 4 -Effect of averaging
323-426	LED	651.9-36	5	30	100	10000	5	t.stage	50	20	N	N	Group 5 - [Cou00]
427-508	Filtered White L. 648.7-12.2	651.9-36	20	30	100	10000	5	t.stage	50	20	N	N	Group 6 - [Cou00]
Rcmin1-4	RCLED MR1051 650nm	648.7-12.2	15	30	100	10000	5	t.stage	50	20	1.5 mm	N	Coherence modulator - variable voltage

File name	target	Filter CWL-FWHM (nm)	average number	amplifier			Sampling (kHz)	Drive method	Scan freq. Hz	Scan (μm)	Iris diam. (mm)	Exit lens	comments
				Gain dB	F_L (Hz)	F_H (Hz)							
5-14	RCLED MR1051	648.7-12.2	5-10	30	100	10000	5	t.stage	50	20	1.5	N	Coherence modulator V=5V
15-24	RCLED MR1051	648.7-12.2	15	30	100	10000	5	t.stage	50	20	4	N	Gr. 7-Coherence modulator V=25V
25-92	RCLED MR1051	648,7-12,2	1	30	100	10000	5	t.stage	50	20	4	Double achromat f=150 mm	Group 8 - RCLED shifts
Lasmin1-80	He-Ne	633-11	1	30	100	10000	10	t.stage	150	20	10	same	Controlled power measurement
81-83	He-Ne	633-11	1	30	100	10000	10	t.stage	150	20	10	same	Group 9 - Fringe combination test
84-213	He-Ne	633-11	4	30	100	10000	10	t.stage	150	20	10	same	Group 10 - Laser shifts
Mono27-34	Laser diode 657-0.1	Monoc. 657-9 nm	5	30	100	10000	5	t.stage	100	40	10	same	Trial of monochromator as a variable filter
35-38	L.diode	Monoc.	1	30	100	10000	1	t.stage	100	200	10	same	Same as above
Minmo1-50	Monoc. 632.6-5.4	633-11	3	30	100	10000	10	t.stage	100	40	10	same	Group 11 -light source at 2500K
51-100	Monoc. 651.9-5.4	651.9-36.2	5	30	100	10000	10	t.stage	100	40	10	same	Group 12 – bandwidth ratio optimisation
101-151	Monoc. 674.8-5.4	674.8-17.8	8	30	100	10000	10	t.stage	100	40	10	same	Group 13 - bandwidth ratio optimisation
152-191	Monoc. 665.9-5.4	674.8-17.8	8	30	100	10000	10	t.stage	100	40	10	same	Group 14 - effect of offset
192-231	Monoc. 670.4-5.4	674.8-17.8	8	30	100	10000	10	t.stage	100	40	10	same	Group 15 - effect of offset

File name	target	Filter CWL-FWHM (nm)	average number	amplifier			Sampling (kHz)	Drive method	Scan freq. Hz	Scan (μm)	Iris diam. (mm)	Exit lens	comments
				Gain dB	F_L (Hz)	F_H (Hz)							
232-272	Monoc. 679.3-5.4	674.8-17.8	5	30	100	10000	10	t.stage	100	40	10	same	Group 16 - effect of offset
273-322	Monoc. 683.7-5.4	674.8-17.8	8	30	100	10000	10	t.stage	100	40	10	same	Group 17 – effect of offset
Min2-5	He-Ne	633-11	16	medium	N	N	2.5 MHz	Fast Piezo	500 Hz (sine)	30	10	same	Fringes 48 kHz - drift assessment
6-115	He-Ne	633-11	16	Med.	N	N	2.5	f. piezo	500	30	10	same	Group 18 – shifts
116-139	He-Ne	633-11	16	Med.	N	N	2.5	f. piezo	500	30	10	same	Minimum detectable shift
140-146	He-Ne	633-11	16	Med.	N	N	1	f. piezo	500	30	10	same	Longer interferograms
147-166	He-Ne	633-11	16	Med.	N	N	2.5	f. piezo	500	30	10	same	Gr. 19 – Gr. 18 continued
167-172	Monoc. 674.8-5.4	674.8-17.8	variable	Med. /high	N	N	2.5/1	f. piezo	500	30	10	same	Monochromator setting-up
173-232	Monoc. 674.8-5.4	674.8-17.8	8	Med.	N	N	2.5	f. piezo	500	30	10	same	Monochromator shifts – alignment problem
233-242	Monoc. 674.8-5.4	674.8-17.8	variable	Med.	N	N	1	f. piezo	500	30	10	same	Monochromator setting-up
243-302	Monoc. 674.8-5.4	674.8-17.8	8	Med/ amp. 30dB	1kHz	1MHz	2.5	f. piezo	500	30	10	same	Group 20 - Monochromator shifts
303-309	Monoc. 674.8-5.4	674.8-17.8	variable	Var.	variable	Var.	variable	f. piezo	500/100	30	10	same	Monochromator setting-up
310-369	Monoc. 674.8-5.4	674.8-17.8	8	Med.	10kHz	100 kHz	1	f. piezo	500	30	10	same	Group 21 - Monochromator shifts

2 – IMAGES

files	target	filter	AGC	gamma	F/#	Iris (mm)	Size (frames)	Lens f (mm)	comments
Min1-27	Monoc. 674.8-5.4	674.8-17.8	on	on	11	17	1	100	Setting-up
28-32	Monoc. 674.8-5.4	674.8-17.8	on	on	11	17	1	100	Stability and drift assessment
33-38	Monoc. 674.8-5.4	674.8-17.8	on	on	11	17	1	100	Setting-up
40-99	Monoc. 674.8-5.4	674.8-17.8	on	on	11	17	1	200	Group 22 – amplitude modulation technique
102-196	He-Ne	633-11	off	off	11	17	1	200	Group 23 – amplitude modulation
Seq1-13	He-Ne	633-11	off	off	11	17	1	200	Setting-up long sequences
Seq14-16	He-Ne	633-11	off	off	11	17	500	200	Group 24 – shifts – test target no.1
Seq17-20	He-Ne	633-11	off	off	11	17	500	200	Group 25 – shifts – test target no.2

Appendix 3 – Movies recorded with the CCD camera in Chapter 6

The files available in the enclosed CD can be viewed with PC-based multimedia players.

File	Comments
three bars.avi	A sequencc of 20 frames, acquired by varying the path difference manually. An extended target, formed by three vertical bars from the USAF 1951 chart, blinks in a white light background.
spot.avi	A sequence of 25 frames, acquired with an automated arrangement. A point target formed by a He-Ne laser spot blinks in a white light background.
fringes1.avi	First section of a sequence of 500 frames, showing the fringe format and the direction of movement of the fringes in the Coherence Imaging Experiment. This movie presents frames 19 to 180. The visibility of the self-coherence function, expressed by the fringe contrast, grows approximately up to frame 130, then decays from this frame onwards. This region is the main lobe of the self-coherence function.
fringes2.avi	Frames 180-279. Fringe visibility decays monotonically.
fringes3.avi	Frames 280-400. The visibility falls approximately until frame 330, which represents the first minimum of the self-coherence function, the region of interest for the IPSS technique. The first sidelobe of the function is shown as an increase in visibility that follows, with the second maximum seen approximately at frame 380.
fringes4.avi	Frames 400-499. The visibility falls indicating the end of the first sidelobe. The second sidelobe does not appear, as its contrast is below the noise level.

Appendix 4 - Conversion of a gas detection sensitivity in ppm.m to an SCR in dB

In his doctoral thesis, Drum [Dru90] performed an experiment to detect the 489 nm spectral line of NO_2 using a 10 cm gas cell, and a 100 W tungsten-halogen lamp as a light source. Using the coherence measurement technique studied in his work, he has reported a sensitivity 15 ppm.m, based on this experiment. However, we have observed in his thesis that this figure is the measurement uncertainty of his experiment. In terms of minimum detectable signal, the smallest concentration-path length measured with error bars above zero was 145 ppm.m. These two figures are considered in the calculations to follow.

Additional data furnished in Drum's thesis are:

- NO_2 peak absorption coefficient at 489 nm: 6.10^{-4} ppm^{-1}. m^{-1};
- Interference filter used in the coherence measurement technique: central wavelength 489 nm, optical bandwidth 7.5 nm in wavelengths, or $\Delta\kappa_B$ =31365 m^{-1} in wavenumbers;
- Width of the 489 nm NO_2 spectral line: 1.5 nm, or $\Delta\kappa_T$ =6273 m^{-1} in wavenumbers.

The SCR, as defined in this thesis, is the ratio between target and background powers:

$$SCR = {}^{P_T}\!\big/_{P_B}$$

The specific spectral absorption coefficient is defined considering the absorption of collimated monochromatic radiation with a spectral density of unity [Can99]. It can be interpreted as the power absorbed in a incremental spectral unit $d\kappa$, where κ is the wavenumber. Considering a background spectral density of 1 W.m, as set by the definition just given, and assuming linear absorption, the target power can be calculated by:

$$P_T = 6.10^{-4}.C.1.\Delta k_T \qquad \text{[W]},$$

where C is the concentration-path length product in ppm.m, and $\Delta\kappa_T$ is the target spectral linewidth in wavenumbers (m^{-1}). The background power, in Watts, can be calculated by $P_B = 1.\Delta\kappa_B$, where $\Delta\kappa_B$ is the background optical bandwidth, given by the interference filter bandwidth, whose spectral shape is assumed to be rectangular. Calculating P_T and P_B using the values of $\Delta\kappa_T$ and $\Delta\kappa_B$ gives:

$$SCR = 1.2.10^{-4}.CL$$

Replacing CL by the two figures given in Drum's thesis, and converting the power ratios to dB, one obtains an SCR of –27.45 dB for a concentration-path length of 15 ppm.m, and an SCR of –17.6 dB for a concentration-path length of 145 ppm.m. For the sake of comparison with IPSS sensitivities, these SCRs are measured after optical filtering.

Appendix 5 - Notation of spectral and coherence variables

The table below shows the various terminologies used in the literature (for example [Hec98], [Bor99] and [Cha79]) in problems involving the Fourier transform relationship between power spectrum and self-coherence function, and was compiled during the course of the work in chapters three and eight. The column "first minimum of coherence function" indicates the path difference for which the sinc-shaped coherence function has its first minimum, assuming a spectrum with a top-hat shape.

Domain	Spectral variable	Self-coherence function (Fourier transform)	First minimum of coherence function	Applications
Spatial	Wavenumber $\kappa=1/\lambda$ [cm^{-1}]	$S(\tau)=\Im\{S(\kappa)\}=\int_{-\infty}^{+\infty}S(\kappa).e^{-j2\pi\kappa\tau}d\kappa$	$\tau=1/\Delta\kappa$ $=\lambda^2/\Delta\lambda$ [μm]	Spectroscopy
Spatial	Propagation number $k=2\pi/\lambda$ [rad/cm]	$S(\tau)=\Im\{S(\kappa)\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}S(\kappa).e^{-j\kappa\tau}d\kappa$	$\tau=2\pi/\Delta\kappa$ $=\lambda^2/\Delta\lambda$ [μm]	Optics
Temporal	Frequency $f=c/\lambda$	$S(\tau)=\Im\{S(f)\}=\int_{-\infty}^{+\infty}S(\kappa).e^{-j2\pi f\tau}df$	$\tau=1/\Delta f$ $=\lambda^2/(c.\Delta\lambda)$ [ps]	Communications theory
Temporal	Angular frequency $\omega=2\pi c/\lambda$ [rad/s]	$S(\tau)=\Im\{S(\omega)\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}S(\omega).e^{-j\omega\tau}d\omega$	$\tau=2\pi/\Delta\omega$ $=\lambda^2/(c.\Delta\lambda)$ [ps]	Communications theory, mathematics

Appendix 6 - Derivation of the self-coherence function of a filtered Lorentzian target and background

Let us assume a Lorentzian curve normalised such that the area below the curve equals one. This is done in order to have a target power P_T, a variable multiplying the Lorentzian. It will be given [Goo85] by:

$$T_U(\kappa) = \frac{2/\pi\delta}{1+(\frac{2}{\delta}(\kappa-\kappa_0))^2} \qquad \text{(A6.1)},$$

where the subscript "u" means unfiltered, δ is the target FWHM in wavenumbers, κ is the wavenumber and κ_0 is the central wavenumber. This spectrum is then passed through a top hat filter centred at κ_0 having a bandwidth, Δκ:

$$T(\kappa) = \frac{2/\pi\delta}{1+(\frac{2}{\delta}(\kappa-\kappa_0))^2}.rect(\frac{|\kappa-\kappa_0|}{\Delta\kappa}) \qquad \text{(A6.2)},$$

where the *rect* function was defined in chapter eight. The self-coherence function of the target can be obtained by applying an inverse Fourier transform (IFT) to the spectrum in eq. A6.2. Using the Fourier Transform displacement theorem, the terms κ-κ_0 produce a phase term $e^{j2\pi\kappa_0\tau}$. Thus the IFT can be expressed as:

$$\Gamma_T(\tau) = \int_{-\Delta\kappa/2}^{+\Delta\kappa/2} \frac{2/\pi\delta}{1+(\frac{2\kappa}{\delta})^2}.e^{j2\pi\tau\kappa}.d\kappa \qquad \text{(A6.3)}$$

A change of variables is now required. Let's make x=2πτκ, then dx=2πτdκ. Replacing these in eq. A6.3 and manipulating gives:

$$\Gamma_T(\tau) = \delta\tau\int\frac{e^{jx}}{(\pi\delta\tau)^2+x^2}dx \qquad \text{(A6.4)}$$

The integral is presented as indefinite in order to allow the usage of a table of integrals. This particular one is tabulated with a=πδτ [Abr72] as:

$$\int\frac{e^{jx}}{a^2+x^2}dx = \frac{j}{2a}\left[e^{-a}.E_1(-a-jx)-e^{a}.E_1(a-jx)\right]+C \qquad \text{(A6.5)},$$

where C is a constant and E_1 is the *exponential integral*, defined as $E_1(z)=\int_z^\infty\frac{e^{-t}}{t}dt$, where z is complex and $|\arg(z)|<\pi$. Substituting eq. A6.5 in eq. A6.4 and the values of *a* and *x* in eq. A6.5 gives:

$$\Gamma_T(\tau) = \frac{j}{2\pi}\left[e^{-\pi\delta\tau}.E_1(-\pi\delta\tau-j2\pi\tau\kappa)-e^{\pi\delta\tau}.E_1(\pi\delta\tau-j2\pi\tau\kappa)\right]$$

Now substituting the integration limits for κ=+/-$\Delta\kappa$/2:

$$\Gamma_T(\tau) = \frac{j}{2\pi}\begin{bmatrix} e^{-\pi\delta\tau}.E_1(-\pi\delta\tau - j\pi\tau\Delta\kappa) - e^{\pi\delta\tau}.E_1(\pi\delta\tau - j\pi\tau\Delta\kappa) - e^{-\pi\delta\tau}.E_1(-\pi\delta\tau + j\pi\tau\Delta\kappa) \\ + e^{\pi\delta\tau}.E_1(\pi\delta\tau + j\pi\tau\Delta\kappa) \end{bmatrix}$$

As the exponential integral has the property $E_1(z^*) = E_1^*(z)$ [Abr72], where the star means complex conjugate, and as the difference between a complex number and its complex conjugate is *2j* times its imaginary part, we can simplify and rewrite the expression as:

$$\Gamma_T(\tau) = \frac{1}{\pi}\left[e^{\pi\delta\tau}.\mathrm{Im}(E_1(\pi\delta\tau - j\pi\tau\Delta\kappa)) - e^{-\pi\delta\tau}.\mathrm{Im}(E_1(-\pi\delta\tau - j\pi\tau\Delta\kappa))\right] \qquad \text{(A6.7)}$$

We can try to analyse eq. A6.7 for two extreme cases. If $\delta >> \Delta\kappa$, that is, if the target bandwidth is much wider than the filter bandwidth, we can neglect the imaginary parts of the arguments of E_1. The exponential integral of a real number is real as well [Abr72], thus $\Gamma_T(\tau)=0$, as we would expect if most of the target energy is lost in a much narrower filter. On the other hand, if $\delta << \Delta\kappa$, or if the target bandwidth is much narrower than the filter bandwidth, we can neglect the real part of the argument of the exponential integral, and regroup eq. A6.7 to produce:

$$\Gamma_T(\tau) = \frac{1}{\pi}(e^{\pi\delta\tau} - e^{-\pi\delta\tau}).\mathrm{Im}(E_1(-\pi\tau\Delta\kappa j)) \qquad \text{(A6.8)}$$

We would expect that this expression would tend to the IFT of the unfiltered Lorentzian in eq. A6.1, which can easily be calculated to be [Hsu84]:

$$\Gamma_{TU}(\tau) = e^{-\pi\delta|\tau|}.e^{-j2\pi\kappa_0\tau} \qquad \text{(A6.9)}$$

It can be seen in eq. A6.8 that its first term, containing the exponentials, diverges for both positive and negative large values of τ. This happens because $\Gamma_T(\tau)$ has a discontinuity of the first derivative at $\tau=0$. Due to this, we have to analyse eq. A6.7 separately for $\tau<0$ and $\tau>0$. For $\tau<0$, $\tau=-|\tau|$, and for $\tau>0$, $\tau=|\tau|$. Thus:

$$\Gamma_T(\tau) = \frac{1}{\pi}\left[e^{-\pi\delta|\tau|}.\mathrm{Im}(E_1(-\pi\delta|\tau| + j\pi|\tau|\Delta\kappa))\Big|_{\tau<0} - e^{-\pi\delta|\tau|}.\mathrm{Im}(E_1(-\pi\delta|\tau| - j\pi|\tau|\Delta\kappa))\Big|_{\tau>0}\right].$$

Using the complex conjugate property of the exponential integral, and the fact that Im(z*)=–Im(z), the second term can be expressed as $+e^{-\pi\delta|\tau|}.\mathrm{Im}(E_1(-\pi\delta|\tau| + j\pi|\tau|\Delta\kappa))\Big|_{\tau>0}$.

Adding the two terms, we finally obtain:

$$\Gamma_T(\tau) = \frac{2}{\pi}\left[e^{-\pi\delta|\tau|}.\mathrm{Im}(E_1(-\pi\delta|\tau| + j\pi|\tau|\Delta\kappa))\right] \qquad \text{(A6.10)}$$

This expression contains the same exponential seen in eq. A6.9. Repeating the analysis with extreme cases, if $\delta << \Delta\kappa$ (narrow target) we can neglect the real part of the argument of the exponential integral. As calculated using the Matlab built-in function *expint*, which implements the exponential integral function, the imaginary part of the exponential integral of an imaginary argument has an oscillatory pattern, with a first null occurring for $\pi\tau\Delta\kappa$=1.925. This gives a path difference given by:

$$\tau = {}^{1.925}\!\big/_{\pi\Delta\kappa} = {}^{0.6127}\!\big/_{\Delta\kappa}\,. \qquad \text{(A6.11)}$$

This is not consistent with the fact that the first null coming from the background with a top hat filter with bandwidth $\Delta\kappa$ occurs at $1/\Delta\kappa$. Another issue is that the filtered expression (eq. A6.10) does not tend to the unfiltered one (eq. A6.9) if the filter is much wider than the target, as one would expect. For verification purposes, the results given by eq. A6.10 were compared with the result of a convolution between a *sinc* function and two exponentials. According to the convolution theorem [Hsu84], the Lorentzian apodised by the top hat in the spectral domain corresponds to the convolution just mentioned in the coherence domain. Fig. A6.1 shows in plot (a) the self-coherence function of a Lorentzian target with a FWHM of 1 nm, filtered by a 11 nm FWHM filter, calculated using eq. A6.10. Plot (b) shows the convolution. The results are similar, although the sidelobes in plot (b) are higher, possibly due to the truncation of the sinc function used in the convolution. The position of the first minimum is slightly different in the two plots, but both around 20 μm. At a wavelength of 633 nm, and using eq. A6.11, the calculated minimum would be at 22 μm, in agreement with plot (a), ensuring that the program is implementing the equation correctly.

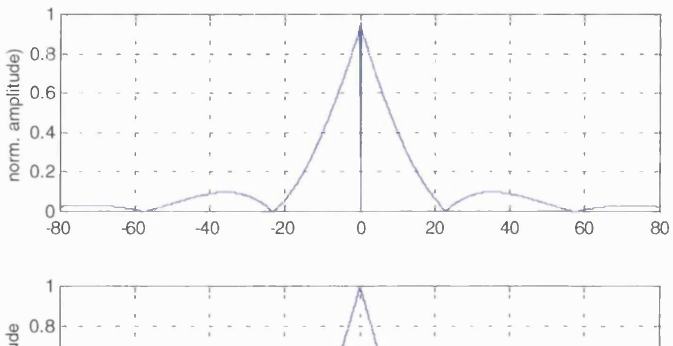


Figure A6.1 - Filtered Lorentzian target. (a) Data calculated using eq. A6.10. (b) - Convolution.

As the theoretical solution for the self-coherence function of a Lorentzian target, considering the effects of filtering, did not agree in the limit with the unfiltered solution, the results of this derivation were not used in the main text of the thesis.

References

Abr72 M. Abramowitz, I.E. Stegun. *Handbook of mathematical functions.* 10th ed., John Wiley and sons, New York, 1972.

Agi01 Agilent 8504B Precision reflectometer user's guide, http://www.agilent.com/cm/rdmfg/pr/

Ari91 A. Arie, M. Tur, E. Goldstein. *Probability-density function of noise at the output of a two-beam interferometer.* J. Opt. Soc. Am. Vol. 8, no. 12, 1991.

Bai94 G.Baister, P.V. Gatenby. *Pointing, acquisition and tracking for optical space communications.* IEE Electronics & communication engineering journal, December 1994.

Bax00 F.O. Baxley, R.B. Sanderson, J.B. Montgomery, J.F. McCalmont. *Evaluation of two-color missile detection algorithms against real backgrounds.* Proceedings of the SPIE vol. 4048: 112-121, 2000.

Ble99 Z. Bleier, C. Brouillette, R. Carangelo. *A monolithic interferometer for FT-IR spectroscopy.* Spectroscopy 14 (10), October 1999.

Bod91 D. Bodson. *Electro-optical communications dictionary.* Hayden Book Company, Rochelle Park, 1991.

Bor99 M. Born, E. Wolf. *Principles of optics - 7th ed.* Cambridge Univ. Press, Cambridge, 1999.

Bra96 E. Branlund, P. Davis, R. Lindgren. *Infrared detection of low-contrast sea-skimming cruise missiles.* Proceedings of the SPIE vol. 2742: 196-207, 1996.

Bre96 J.B. Breckinridge. *Evolution of imaging spectrometry: past, present and future.* Proceedings of the SPIE vol. 2819: 2-6, 1996.

Bur98 C.S. Burrus, R.A Gopinath, H. Guo. *Introduction to wavelets and wavelet transforms: a primer.* Prentice-Hall, New Jersey, 1998.

Cam93 S.B. Campana, editor. *Passive electro-optical systems.* Vol. 5 of The Infrared & Electro-Optical Systems Handbook. ERIM/SPIE Press, Bellingham, 1993.

Can99 J. Caniou. *Passive infrared detection.* Kluwer Academic Publishers, Boston, 1999.

Cha79 J. E. Chamberlain. *The principles of interferometric spectroscopy.* Wiley-Interscience, New York, 1979.

Che96 S. Chern, K. Chao. *An enhanced Fourier transform spectrometer with a search algorithm.* IEEE Transactions on Instrumentation and Measurement, vol. 45, no.1: 127-135, 1996.

Che97 J. Cheung, D. Ferris, L. Kurz. *On classification of multispectral infrared image data.* IEEE Transactions on Image Processing, vol. 6, no. 10: 1456-1459, 1997.

Chu01 H. Chun, D. Ma, W. Li. *Thin-film filters are the building blocks of multiplexing devices.* WDM Solutions, May 2001.

Cor88 A. Corney. *Atomic and laser spectroscopy.* Oxford University Press, New York, 1988.

Cou00 R.C. Coutinho,D.R. Selviah, H.A. French. *Detection of Partially Coherent Optical Emission Sources.* Optical Pattern Recognition XI, David P. Casasent, Tien-Hsin Chao, Editors, Proceedings of the SPIE Vol. 4043, pp. 238-248, 2000.

Cou02 R.C. Coutinho, D.R. Selviah, R.F. Oulton, J.W. Gray, P.N. Stavrinou, H.D. Griffiths, G. Parry. *Variable Numerical Aperture Temporal Coherence Measurement of Resonant Cavity LEDs.* Journal of Lightwave Technology, accepted for publication.

Cou99 R.C. Coutinho, H.A. French, D.R.Selviah,D. Wickramasinghe,H.D. Griffiths. *Detection of Coherent Light in an Incoherent Background.* Proceedings of the 12th Annual Meeting of the IEEE Lasers and Electro-Optical Society, San Francisco, USA, vol.1, pp. 247-248, 1999.

Der96 E.L. Dereniak, G.D. Boreman. *Infrared detectors and systems.* John Wiley & Sons, New York, 1996.

Der99 Interview with Duleep Wickramasinghe, Defence and Evaluation Research Agency, Portwsdown West, 1999.

Des96 M.R. Descour. *The throughput advantage in imaging Fourier-transform spectrometers.* Proceedings of the SPIE vol. 2819: 285-290, 1996.

Dia01 M. Diani, A. Baldaci, G. Corsini. *Novel background removal algorithm for Navy infrared search and track systems.* Optical Engineering vol. 40, no. 8: 1729-1734, 2001.

Dru87 S.M. Drum. *Remote gas detection by coherence measurement.* Proceedings of the SPIE vol. 807:86-90, 1987.

Dru89 S.M. Drum. *Remote sensing of gases by the analysis of temporal coherence.* Proceedings of the SPIE vol. 1129: 52-59, 1989.

Dru90 S.M. Drum. *The remote detection of gases using coherence measurement.* PhD thesis, University of Southampton, 1990.

Duf89 C.J. Duffy, D. Hickman. *A temporal coherence-based optical sensor.* Sensors and Actuators 18: 17-31, 1989.

Eis96 M.T. Eismann, C.R. Schwarz, J.N. Cederquist, J.A. Hackwell, R.J. Huppi. *Comparison of infrared imaging hyperspectral sensors for military target detection applications.* Proceedings of the SPIE vol. 2819: 91-101, 1996.

Eis97 M.T. Eismann, C.R. Schwarz. *Focal plane array nonlinearity and nonuniformity impacts to target detection with thermal infrared imaging spectrometers.* Proceedings of the SPIE vol. 3063: 164-173, 1997.

Elw00 C. Elwell, J. Hebden. *Near-infrared spectroscopy.* University College London, Biomedical Optics Research Group web site, http://www.medphys.ucl.ac.uk/research/borg/research/NIR_topics/nirs.htm

ESA01 *A world first: data transmission between European satellites using laser light.* European Space Agency web-site, http://esa.int/export/esaCP/ESASGBZ84UC_Improving_0.html, September 2002.
Fel84 P.B. Felgett. *Three concepts make a million points.* Infrared Phys. Vol. 24, no. 2/3: 95-98, 1984.
Fre72 H.A. French. *A new electro-optical technique employing modulation of Fourier transform images.* ASWE Report XAR(S)/250/1972.
Fre84 H.A. French. *A new pre-detector processing technique for optical remote sensing systems.* Proceedings of the 1984 International Symposium on Noise and Clutter Rejection in Radars and Imaging Sensors: 188-192,1984.
Fre84A H.A. French. *Apparatus and method for remote sensing of gases, vapors or aerosols.* US Patent No. 4,676,642, 1987 (Filed in1985)
Fre88 H.A. French. *Introduction to optical transform image modulation techniques and their application to ESM data processing.* IEE Colloquium on Signal Processing for ESM Systems:9/1-9/2, 1988.
Fre93 H.A. French, P.Sutton. *Electrical signal event marker circuit.* UK patent no. 2261727A, Farnborough, 1993.
Gar92 W.A. Gardner. *A unifying view of coherence in signal processing.* Signal processing vol.29:113-140, 1992.
Goe95 A.F. Goetz. *Imaging spectrometry for remote sensing: vision to reality in 15 years.* Proceedings of the SPIE vol. 2480: 2-13, 1995.
Goo85 J.W. Goodman. *Statistical Optics.* John Wiley and sons. New York, 1985.
Gra01 J.W. Gray, R.F. Oulton, P.N. Stavrinou, G. Parry, G. Duggan, R.C. Coutinho, D.R. Selviah. *Angular Emission Profiles and Coherence Length Measurements of Highly Efficient, Low Voltage Resonant Cavity Light Emitting Diodes Operating Around 650nm.* Proceedings of the SPIE vol. 4278:81-89 , 2001.
Hec98 E. Hecht. *Optics - 3rd edition.* Addison-Wesley, New York, 1998.
Hic88 D. Hickman, C.J. Duffy. *Incorporation of imaging into a temporal coherence sensor.* Optics Letters vol 13, no. 10:898-900, 1988.
Hic89 D. Hickman. *The selective detection and analysis of weak signals in optically cluttered images using temporal coherence information.* PhD thesis, King's College, 1989.
Hil00 C.I. Hilliard. *Selection of a clutter rejection algorithm for real-time target detection from an airborne platform.* Proceedings of the SPIE vol. 4048: 74-84, 2000.
Hit99 C.K. Hitzenberger, A. Baumgartner, H. Sattman, A.F. Fercher. *Differential phase measurements by partial coherence interferometry.* Proceedings of the SPIE vol. 3598:145-151, 1999.
Hor96 R.F. Horton. *Optical design for a high étendue imaging Fourier transform spectrometer.* Proceedings of the SPIE vol. 2819: 300-315, 1996.

Hot00 K. Hotate, Z. He, T. Kitazume. *Surface shape measurement for a multilayered object by synthesis of optical coherence function.* IEEE Journal on Selected Topics in Quantum Electronics, vol. 6, no. 5: 723-728, 2000.

Hot01 K. Hotate, M. Tanaka. *Making sense: Correlation-based continuous-wave techniques provide high spatial resolution for distributed fiber-optic sensing.* SPIE's OE Magazine, November 2001:36-40, 2001.

Hsu84 H.P. Hsu. *Applied Fourier analysis.* Harcourt Brace College publishers, New York, 1984.

Hud69 R.D. Hudson. *Infrared system engineering.* Wiley-Interscience, New York, 1969.

Ico01 *Resonant Cavity LED's*, http://www.ee.ic.ac.uk/ElecMatDev/

Iva00 E.V. Ivanov. *Static Fourier transform spectroscopy with enhanced resolving power.* J. Opt. A: Pure Appl. Opt. Vol. 2: 519-528,2000.

Kay88 S.M. Kay. *Modern spectral estimation.* Prentice-Hall, Englewood Cliffs, 1988.

Kim99 H.S. Kim, R. P. Haaksman, T.P. Newson, D.J. Richardson. *Noise properties and phase resolution of interferometer systems interrogated by narrowband fiber ASE sources.* Journal of Lightwave Technology, vol. 17, no.11:2327-2334, 1999.

Kre88 E. Kreyszig. *Advanced engineering mathematics.* John Wiley and sons, New York, 1988.

Lju97 P. Ljungberg, R.G. Kihlén, S.H. Lundqvist, P. Potet, S. Berrebi. *Multispectral imaging MWIR sensor for determination of spectral target signatures.*Proceedings of the SPIE vol. 3061:823-832, 1997.

Lob01 M.I. Lobachev, D.V. Lyakin, V.P. Ryabukho. *Peculiarities of interferometer signal in partially coherent light of multiple source.* Proceedings of the SPIE vol. 4242:237-242, 2001.

Man65 L. Mandel, E. Wolf. *Coherence properties of optical fields.* Rev. Mod. Phys., vol. 37, no. 2: 231-283, 1965.

Man96 V.A. Manasson, L.S. Sadovnik. *Laser warning receiver based on coherence discrimination.* IEEE Proceedings of the National Aerospace and Electronics Conference vol 2: 869-873, 1996.

Mar99 W.J. Marinelli, C.M. Gittins, A.H. Gelb, B.D. Green. *Tunable Fabry-Perot etalon-based long-wavelength infrared imaging spectroradiometer.* Applied Optics vol. 38, no. 12: 1999.

McC96 K. J. McCamey, R. Marko, L. Garvin, T. Peli. *Detection in clutter enhancement.* Proceedings of the SPIE vol. 2742: 225-235, 1996.

McD98 R.N. McDonough, A.D. Whalen. *Detection of signals in noise.* Academic Press, New York, 1998.

MCT02 Military Critical Technologies web-site, http://www.dtic.mil/mctl/ , September 2002.

MeG00 D. MacDonald. Melles Griot. Private communication, 13/10/99.

MeG99 Melles Griot product catalogue, 1999.

Mel99 Melles Griot product catalogue. Helium neon lasers data sheet, pp. 522, 1999.

Mer97 S.P. Merlo, R. G. Lindgren, P.J. Davis. *Application of bayesian field track-before-detect and peak likelihood track-after-detect trackers to shipboard infrared search and track.* Proceedings of the SPIE vol. 3163: 194-203, 1997.

Mon00 J.B. Montgomery, R.B. Sanderson, F.O. Baxley. *Tactical mid-infrared background suppression in heavy clutter environments.* Proceedings of the SPIE vol. 4048: 23-34, 2000.

Mon96 J.B. Montgomery, R.B. Sanderson, F.O. Baxley. *Two color missile signature measurements.* IEEE Proceedings of the National Aerospace Electronics Conference, Dayton, Ohio, USA: 782-786, 1996.

Mou02 J.R. Moulton, C.E. Fink. *Ray tracing approach for realistic hyperspectral forest canopies.* Proceedings of the SPIE vol. 4718, pp.46-64, 2002.

Moy00 Interview with Dr. Tony Moy, Chief Engineer, EO Systems, BAE Systems Plymouth, 13/07/2000.

Nee02 F.P. Neele, W. de Jong, *Prewetting systems as na IR signature control tool.* Proceedings of the SPIE vol. 4718, pp.156-163, 2002.

Nel00 C. Nelsson, F. Lundberg, P. Nilsson, M. Berglund. *Scanspec: an imaging FTIR spectrometer.* Proceedings of the SPIE vol. 4029: 324-336, 2000.

Nic95 T.W. Nicholls, G.D. Boreman, J.C. Dainty. *Use of a Shack-Hartmann wave-front sensor to measure deviations from a Kolmogorov phase spectrum.* Optics Letters vol. 20, no. 24, 1995.

OLE98 *Light absorption indicates best time for cancer therapy.* Optics and Laser Europe, pp. 47-48, April 1998.

Ope89 A.V. Oppenheim, R.W. Schafer. *Discrete-Time Signal Processing.* Prentice Hall, Englewood Cliffs, 1989.

Ori00 Oriel Instruments. *The Book of Photon Tools.*Product catalogue, Stratford, USA, 2000.

Oul01 R.F. Oulton, J.W. Gray, P.N. Stavrinou, G. Parry. *Insight into planar microcavity emission as a function of numerical aperture.* Optics Communications 195: 327-338, 2001.

Pap91 A. Papoulis. *Probability, random variables, and stochastic processes.* 3rd ed. McGraw-Hill, New York, 1991.

Pel97 T. Peli, P. Monsen, R. Stahl, M. Pauli, K. McCamey. *Signal processing improvements for infrared missile warning sensors.* Proceedings IEEE 1997, vol. 2: 1052-1064, 1997.

Per95 M.J. Persky. *A review of spaceborne infrared Fourier transform spectrometers for remote sensing.* Review of Scientific Instruments vol. 66, no. 10, 1995.

Pet90 V.G. Peters, D. R. Wyman, M.S. Patterson, G.L. Frank. *Optical properties of normal and diseased human breast tissues in the visible and near infrared.* Phys. Med. Biol, vol. 35, no. 9, pp. 1317-1334, 1990.

Pis78 N. Piskunov. *Calculo diferencial e integral.* 4th edition (spanish). Editora Mir, Moscow, 1978.

Pol00 Interview with Dr. John Pollard, Senior Scientist, U.S. Army Night Vision and Electro-Optical Systems Directorate, 06/07/2000

Pol93 D.H. Pollock, editor. *Countermeasure systems.* Vol. 7 of The Infrared & Electro-Optical Systems Handbook. ERIM/SPIE Press, Bellingham, 1993.

Pre92 W. Press. *Numerical recipes in C.* 2nd edition, Cambridge University Press, New York, 1992.

Pri97 A.T. Pritt, Jr., P.N. Kupferman, S.J. Young, R.A. Keller. *Imaging LWIR spectrometers for remote sensing applications.* Proceedings of the SPIE vol 3063: 138-149, 1997.

Pro96 J.G. Proakis, D.G. Manolakis. *Digital signal processing.* 3rd ed., Prentice Hall, New Jersey, 1996.

Rap02 J.L. Rapanotti, B. Gilbert, G. Richer, R. Stowe. *IR sensor design insight from missile-plume prediction models.* Proceedings of the SPIE vol. 4718, pp.289-300, 2002.

Rei99 S.M. Reiss. *Optical coherence tomography: an imaging modality for the new millenium.* Biophotonics International, October 1999:36-45, 1999.

San96 R.B. Sanderson. *Infrared missile warning sensors. IEEE* Proceedings of the National Aerospace Electronics Conference, Dayton, Ohio, USA: 777-781, 1996.

Sch96 C.R. Schwartz, M.T. Eismann, J.N. Cederquist. *Thermal multispectral detection of military vehicles in vegetated and desert backgrounds.* Proceedings of the SPIE vol. 2742: 286-297, 1996.

Sch99 J.M. Schmitt. *Optical coherence tomography (OCT): a review.* IEEE Journal on Selected Topics in Quantum Electronics, vol. 5, no. 4: 1205-1215, 1999.

Sel02 D.R. Selviah, E. Stamos. *Similarity suppression algorithm for designing pattern discrimination filters.* Asian Journal of Physics, vol 11, no 2, pp. 1-12, 2002.

She99 L.M. Sheppard. *Novel LEDs could eliminate bacteria.* Photonics Spectra, July 1999.

Shn00 P. Shnitser, I. Agurok, S. Sandomirsky, A. Avakian, R. Lockwood. *Real-time polarization sensitive multispectral automatic imaging system for object contrast enhancement and clutter mitigation.* Proceedings of the SPIE vol. 4025: 229-237, 2000.

Sin00 P.F. Singer, H.A. Quevedo. *Multispectral track features for the general case of an unknown target spectral signature.* Proceedings of the SPIE vol. 4048: 558-567, 2000.

Sin98 P.F. Singer. *A performance model for unresolved target detection using multispectral infrared data.* Proceedings of the SPIE vol. 3373: 16-23, 1998.

Sin99 P.F. Singer. *Fully adaptive space-spectral detection of small targets in the absence of a priori knowledge of the spectral signature of the target.* Proceedings of the SPIE vol. 3809: 131-140, 1999.

Sko90 M.I. Skolnik. *Radar handbook.* McGraw Hill, New York, 1990.

Smi93 F.G. Smith, editor. *Atmospheric propagation of radiation.* Vol. 2 of The Infrared & Electro-Optical Systems Handbook. ERIM/SPIE Press, Bellingham, 1993.

Sta99 P.N. Stavrinou, M. Whitehead, G. Parry, C.C. Button. *Angular spectrum of visible resonant cavity light-emitting diodes.* J. Appl. Phys. 86 (6): 3475-3477, 1999.

Sut82 P. Sutton. *A novel optical pre-detector signal processing technique.* PhD thesis, University of London, 1982.

Sut84 P. Sutton. *Signal and background discrimination using optical domain processing.* Proceedings of the 1984 International Symposium on Noise and Clutter Rejection in Radars and Imaging Sensors: 193-198,1984.

Sut89 P.Sutton. *A novel electro-optical remote-sensing technique based on bandpass coherence processing.* J.Phys.D: Appl. Phys.22:379-384, 1989.

Swa93 N.L. Swanson, J.D. Kalata. *Measurement of temporal coherence loss in laser light scattered from simulated coastal waters.* Proceedings of the IEEE Southeastcon' 93 conference, 1993.

Swa93a N.L. Swanson, D.H. VanWinkle. T*emporal coherence loss due to dynamic scattering of laser light.* Proceedings of the IEEE Oceans' 93 conference, vol 3, pp. III/171-III-175, 1993.

Tar00 A. Tartarovsky, R. Blazek. *Effective adaptive spatial-temporal technique for clutter rejection in IRST.* Proceedings of the SPIE vol. 4048: 85-95, 2000.

Tar99 A. Tartarovsky, S. Kligys, A. Petrov. *Adaptive sequential algorithms for detecting targets in a heavy IR clutter.* Proceedings of the SPIE vol. 3809: 119-130, 1999.

Thé01 J. Thériault. *Passive monitoring of chemical vapor concentration: method and results for ammonia.* Proceedings of the SPIE vol. 4168: 211-221, 2001.

Tol98 T. Tolker-Nielsen, J-C. Guillen, *SILEX: The first European optical communication terminal in orbit. European Space Agency bulletin*, November 1998.

Ton96 S.M. Tonissen, R.J. Evans. *Performance of dynamic programming techniques for track-before-detect.* IEEE Transactions on Aerospace and Electronic Systems, vol. 32, no. 4:1440-1451, 1996.

Upt97 G. Upton, I. Cook. *Understanding statistics.* Oxford university press, Oxford, 1997.

Van68 H. L. Van Trees. *Detection, estimation, and modulation theory.* John Wiley & Sons, New York, 1968.

Wee98 A.R. Weeks. *Fundamentals of electronic image processing.* SPIE/IEEE Press, New York, 1998.

Wei95 P. Wei, J. Zeidler, W. Ku. *Analysis of multiframe target detection using pixel statistics.* IEEE Transactions on Aerospace and Electronic Systems vol. 31, no.1: 238-247, 1995.

Wol97 W.L. Wolfe. *Introduction to imaging spectrometers.* SPIE Press, Bellingham, 1997.

Zis93 G. J. Zissis, editor. *Sources of radiation.* Vol. 1 of The Infrared & Electro-Optical Systems Handbook. ERIM/SPIE Press, Bellingham, 1993.