TIME-FREQUENCY ANALYSIS OF NATIVE AND PROSTHETIC HEART VALVE SOUNDS

Paul Bentley

A thesis submitted for the degree of Doctor of Philosophy
The University of Edinburgh
- June 1996 -
Abstract

In the past, a number of researchers have applied various spectral estimation techniques in an attempt to analyse recorded heart sounds. The majority of these studies have used spectral estimation algorithms such as the Fourier transform and various autoregressive modelling techniques. Despite the definite potential these techniques have shown for the diagnosis of valvular heart disease, they are limited by their assumption of signal stationarity and lack of relation to present stethoscope-based medical evaluation procedures. A solution to these limitations can be achieved by analysing the recorded sounds in the time-frequency domain rather than in the frequency-domain or time-domain independently.

The research detailed in this thesis investigates the application of time-frequency techniques to the description and analysis of recorded heart sounds. Time-frequency is further investigated as a tool for the description of heart sounds in an attempt to diagnose valvular heart disease.

Data used in the study was recorded from 100 subjects in four main valve populations. The four populations investigated were subjects with native heart valves, Carpentier-Edwards bioprosthetic heart valves, Bjork-Shiley metallic prosthetic heart valves and subjects before and after surgery for heart valve replacement.

Prior to the analysis of these data sets, an investigation was performed into the suitability of various time-frequency techniques to the analysis of heart sounds. By comparing the short-time Fourier transform, wavelet transform, Wigner distribution and the Choi-Williams distribution it was found that the Choi-Williams distribution provides definite advantages over the other techniques due to its high resolution and reduced interference properties. Applying the Choi-Williams distribution to typical examples of each data set demonstrated that time-frequency offers definite potential as a heart sound descriptor. Typical results also demonstrate that time-frequency can be used as an aid to understanding the origins of heart sounds.

The work is concluded with the development of a classification scheme designed to diagnose valve condition in the native and bioprosthetic valve populations. Classification was performed using morphological features extracted using the Choi-Williams distribution and via an optimised feature set extracted using the discrete wavelet transform. Classification results suggest that the time-frequency analysis of recorded heart sounds can be used successfully for the identification of valve condition and position of dysfunction.
I would like to thank my supervisors, Dr. Edward McDonnell and Professor Peter Grant, for their support and guidance throughout the course of this work.

Thanks go to my colleagues in the Signals and Systems Group, in particular, my friends Mr. Herkole Sava and Dr. Rajan Bedi for their contribution as partners in the development of equipment and recording of data used in the work.

I am grateful to the staff at the Royal Infirmary of Edinburgh and the Astley Ainslie Hospital Edinburgh for providing access to patients and for their co-operation during recording sessions.

I also acknowledge the financial support provided by EPSRC during the period of my study.

Finally, I would like to thank Miss. Alison Holland and Mr. John Holland for their contribution to this thesis and related publications as a first stage of review. I would also like to thank Miss. Alison Holland for her continuous support during my period of study.
Contents

Abstract

Declaration of Originality

Acknowledgements

Contents

Abbreviations

Glossary of Medical Terms

1 Introduction

1.1 Objectives and Motivation

1.1.1 Auscultation

1.1.2 The uncertainty of origin

1.1.3 Heart valve degeneration

1.2 Biomedical Signal Processing

1.3 Contents of the Thesis

2 Introduction to Heart Sound Signal Analysis

2.1 Heart Sounds

2.1.1 Circulation

2.1.2 Sound origins

2.1.3 Discussion

2.2 The Dysfunctioning Case

2.2.1 Valvular heart disease

2.2.2 Treatment of valvular heart disease

2.3 Introduction to Time-Frequency Signal Description

2.4 Time-Frequency and Spectral Estimation of Heart Sounds: Past Work

2.5 Summary

3 Data Acquisition

3.1 Data Acquisition Hardware

3.1.1 Portable computer / ADC card

3.1.2 Analogue pre-processing

3.2 Phonocardiographic Transducer

3.3 Recording Protocol

3.4 Patient Population

3.5 Typical Results

3.6 Summary
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Time-Frequency Spectral Estimation</td>
<td>44</td>
</tr>
<tr>
<td>4.1</td>
<td>The Short-Time Fourier Transform (STFT)</td>
<td>44</td>
</tr>
<tr>
<td>4.2</td>
<td>The Wavelet Transform (WT)</td>
<td>50</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Implementation of the WT</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>The Wigner Distribution (WD)</td>
<td>62</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Strengths and weaknesses of the WD</td>
<td>66</td>
</tr>
<tr>
<td>4.4</td>
<td>The Choi-Williams Distribution (CWD)</td>
<td>67</td>
</tr>
<tr>
<td>4.5</td>
<td>Data Display Tool</td>
<td>75</td>
</tr>
<tr>
<td>4.6</td>
<td>Comparative Analysis of Time-Frequency Techniques</td>
<td>76</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Time-Frequency analysis of a three sinusoid step signal</td>
<td>77</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Time-Frequency analysis of a dual component LFM test signal</td>
<td>79</td>
</tr>
<tr>
<td>4.6.3</td>
<td>Time-Frequency analysis of a modelled heart sound</td>
<td>83</td>
</tr>
<tr>
<td>4.7</td>
<td>Summary</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>Data Analysis</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>Time-Domain Sound Pre-Processing and Selection</td>
<td>90</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Typical sound extraction</td>
<td>91</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Extraction of time information from the ECG</td>
<td>93</td>
</tr>
<tr>
<td>5.2</td>
<td>Time-Frequency Analysis of Native Valve Sound Data</td>
<td>94</td>
</tr>
<tr>
<td>5.3</td>
<td>Time-Frequency Analysis of Bioprosthetic Valve Sound Data</td>
<td>103</td>
</tr>
<tr>
<td>5.4</td>
<td>Time-Frequency Analysis of Mechanical Valve Sound Data</td>
<td>109</td>
</tr>
<tr>
<td>5.5</td>
<td>Time-Frequency Analysis of Pre &amp; Post Operative Valve Sound Data</td>
<td>116</td>
</tr>
<tr>
<td>5.6</td>
<td>Summary</td>
<td>122</td>
</tr>
<tr>
<td>6</td>
<td>Feature Extraction and Classification</td>
<td>126</td>
</tr>
<tr>
<td>6.1</td>
<td>Feature Extraction</td>
<td>126</td>
</tr>
<tr>
<td>6.2</td>
<td>Morphological Feature Classification</td>
<td>127</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Classical diagnostic parameters and morphological features</td>
<td>128</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Overall classification procedure</td>
<td>130</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Binary feature classification</td>
<td>132</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Continuous feature classification</td>
<td>133</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Classification of native valve subjects using morphological features</td>
<td>142</td>
</tr>
<tr>
<td>6.2.6</td>
<td>Classification of aortic C-E bioprosthetic valve subjects using morphological features</td>
<td>146</td>
</tr>
<tr>
<td>6.3</td>
<td>Optimised Feature Classification</td>
<td>147</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Feature subset selection and classification of native valve patient data</td>
<td>151</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Feature subset selection and classification of aortic C-E bioprosthetic valve patient data</td>
<td>155</td>
</tr>
<tr>
<td>6.4</td>
<td>Summary</td>
<td>158</td>
</tr>
<tr>
<td>7</td>
<td>Conclusions</td>
<td>161</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary and Conclusions of the Research</td>
<td>161</td>
</tr>
<tr>
<td>7.2</td>
<td>Achievements</td>
<td>165</td>
</tr>
<tr>
<td>7.3</td>
<td>Further Work</td>
<td>166</td>
</tr>
<tr>
<td>A</td>
<td>List of Publications</td>
<td>180</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>analogue to digital converter</td>
</tr>
<tr>
<td>AR</td>
<td>autoregressive</td>
</tr>
<tr>
<td>C-E</td>
<td>Carpentier-Edwards</td>
</tr>
<tr>
<td>CWD</td>
<td>Choi-Williams distribution</td>
</tr>
<tr>
<td>CWT</td>
<td>continuous wavelet transform</td>
</tr>
<tr>
<td>DWT</td>
<td>discrete wavelet transform</td>
</tr>
<tr>
<td>ECG</td>
<td>electrocardiogram</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>IFFT</td>
<td>inverse fast Fourier transform</td>
</tr>
<tr>
<td>LFM</td>
<td>linear frequency modulation</td>
</tr>
<tr>
<td>PCG</td>
<td>phonocardiogram</td>
</tr>
<tr>
<td>QMF</td>
<td>quadrature mirror filter</td>
</tr>
<tr>
<td>RID</td>
<td>reduced interference distribution</td>
</tr>
<tr>
<td>RIE</td>
<td>Royal Infirmary of Edinburgh</td>
</tr>
<tr>
<td>STFT</td>
<td>short-time Fourier transform</td>
</tr>
<tr>
<td>WD</td>
<td>Wigner distribution</td>
</tr>
<tr>
<td>WT</td>
<td>wavelet transform</td>
</tr>
</tbody>
</table>
# Glossary of Medical Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aortic heart valve</td>
<td>A semilunar valve that connects the heart’s left ventricle to the arterial network.</td>
</tr>
<tr>
<td>Atrio-ventricular valves</td>
<td>The mitral and tricuspid valves that connect the atrium and ventricles of the heart.</td>
</tr>
<tr>
<td>Atrium</td>
<td>The upper chambers of the heart which receive blood returning from the body via the veins.</td>
</tr>
<tr>
<td>Auscultation</td>
<td>The act of listening to the sounds and murmurs produced by the operation of the heart using a stethoscope.</td>
</tr>
<tr>
<td>Cardiology</td>
<td>The term applied to the branch of medical science devoted to the study of the diseases of the heart.</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>Pertaining to or comprising the heart and its connected blood vessels.</td>
</tr>
<tr>
<td>Chordae</td>
<td>Tendons connected to the heart valves. The heart strings.</td>
</tr>
<tr>
<td>Diastole</td>
<td>The relaxation period of the heart’s cycle during which blood returns to the heart from the body.</td>
</tr>
<tr>
<td>Electrocardiogram</td>
<td>A record of the variations in electrical potential which occur in the heart as it contracts and relaxes.</td>
</tr>
<tr>
<td>Heart sounds</td>
<td>Used generally as a term for the collection of sounds that are assumed to be a result of the action of the heart.</td>
</tr>
<tr>
<td>Hemodynamics</td>
<td>The circulation of blood.</td>
</tr>
<tr>
<td>Intercostal space</td>
<td>A term applied to the gaps between the ribs.</td>
</tr>
<tr>
<td>In-vitro</td>
<td>Literally &quot;in a glass&quot;, referring to observations made outside the body.</td>
</tr>
<tr>
<td>In-vivo</td>
<td>The opposite of in-vitro and referring to observations of processes in the body.</td>
</tr>
<tr>
<td>Mitral heart valve</td>
<td>A bicuspid atrio-ventricular valve on the left side of the heart.</td>
</tr>
<tr>
<td>Murmur</td>
<td>The uneven, rustling sounds heard by auscultation over the heart and various blood-vessels in abnormal conditions.</td>
</tr>
<tr>
<td><strong>Myocardium</strong></td>
<td>The muscular substance of the heart.</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td><strong>Phonocardiogram</strong></td>
<td>The graphical recording of heart sounds and murmurs.</td>
</tr>
<tr>
<td><strong>Prosthesis</strong></td>
<td>An artificial replacement for a human organ.</td>
</tr>
<tr>
<td><strong>Pulmonary heart valve</strong></td>
<td>A semilunar valve that connects the heart's right ventricle to the arterial network.</td>
</tr>
<tr>
<td><strong>Regurgitation</strong></td>
<td>Applied, in this work, to the return of blood across an insufficient closed valve.</td>
</tr>
<tr>
<td><strong>Stenosis</strong></td>
<td>A term applied to the condition of unnatural narrowing in a passage or orifice of the body. The word is especially used in connection with the four openings of the heart at which the valves are situated.</td>
</tr>
<tr>
<td><strong>Sternal</strong></td>
<td>Pertaining to the sternum, commonly referred to as the breast bone.</td>
</tr>
<tr>
<td><strong>Systole</strong></td>
<td>The contraction of the heart that alternates with the resting phase, known as diastole.</td>
</tr>
<tr>
<td><strong>Tricuspid heart valve</strong></td>
<td>An atrio-ventricular valve on the right side of the heart.</td>
</tr>
<tr>
<td><strong>Thrombosis</strong></td>
<td>Formation of a blood-clot within the vessels of the heart.</td>
</tr>
<tr>
<td><strong>Valvular heart disease</strong></td>
<td>Degeneration or dysfunction of the valves of the heart.</td>
</tr>
<tr>
<td><strong>Ventricle</strong></td>
<td>One of the two muscular lower chambers of the heart.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Prior to 1818, the only methods available for the examination of the chest were feeling with the hand and immediate auscultation with the ear directly on the chest. A physician of the time, R.T.H. Laennec, described the objections to putting the ear directly on the chest:

"It is always inconvenient, both to the physician and the patient. In the case of females it is not only indelicate, but often impractical and for the class of person found in the hospital it is disgusting."

At this time only charity patients went to hospital, doctors made house calls to their paying customers.

Laennec used immediate auscultation until 1816 when he was examining a girl with general symptoms of a diseased heart. Due to the fact that she was obese, young and female he felt the usual examination methods were inappropriate. He immediately rolled several pieces of paper into a cylinder and held one end to his ear and the other to the girl's chest above her heart. The results were dramatic and encouraged Laennec to improve this instrument. Eventually he developed a hollow wooden cylinder 30 cm long with an inner diameter of approximately 1 cm and an outer diameter of about 7.5 cm. He named it the stethoscope.

Since its invention, the stethoscope has proved itself probably the most useful tool in medicine. The measure of its success is that it is almost universally accepted as the symbol of medicine and the physician. The work described in this thesis is an attempt to analyse the potential of modern data capture and signal processing techniques, as an alternative to the stethoscope and the human listener.
Chapter 1: Introduction

The following sections of the introduction describe the motivation behind the work, including reasons for attempting the work and the evidence available that points to the validity of the investigation. Consideration is given to the unique problems associated with the analysis of biomedical signals and finally, a chapter by chapter description is given of the work presented in this thesis.

1.1 Objectives and Motivation

The aim of the work in this thesis was to investigate the suitability of modern time-frequency techniques for the analysis of surface-recorded heart sounds. The specific objectives set to achieve this aim were:

- To collect a representative data set of subjects from a number of population groups. The collection of a substantial data set allows the presentation of a number of comparative results which in turn allow clear conclusions to be made.
- To compare the suitability of a number of time-frequency techniques for the analysis and description of heart sounds.
- To time-frequency analyse the data set and subsequently investigate the effect that valve type, condition and construction has on the appearance of the time-frequency sound description.
- To investigate the use of time-frequency methods as feature extraction techniques for the automatic evaluation of valve condition. By investigating the validity of such an automated scheme it is envisaged that the potential of time-frequency as an informative tool will be demonstrated.

The following three sections describe the various motivations that have lead to the consideration of time-frequency methods as a possible aid to heart sound analysis.

1.1.1 Auscultation

Auscultation is commonly defined as the act of listening to the sounds and murmurs produced by the heart using a stethoscope. There are three major components to the auscultation process: the collection of sounds using the stethoscope, the detection of sounds via the auditory system and the recognition and diagnosis of condition that occurs in the listener's brain. The major limitation of the auscultation process occurs in the human auditory system as
collected sounds are being detected. The human auditory system is particularly inefficient when low frequency, low intensity sounds are being analysed [1]. In an attempt to overcome the limitations of the auditory system early work detailed the electronic recording of sounds to generate a moving paper record. For a period of time these paper phonocardiograms (PCGs) were used in a number of cardiology units as aids to the diagnosis of pathological heart conditions. In spite of the increase in available information that the PCG provides, the PCG has now almost completely disappeared as a diagnostic aid and can often only be found being used as a means of illustration in literature. The reason given for this abandonment of the PCG is that in certain situations it does not show sounds that an experienced auscultator can hear [2]. The reason for this inconsistency is quite probably the alternative descriptions provided by that the PCG and the auditory system. As described by Gabor [3], the human auditory system provides a description of sound in a joint time and frequency format. The motivation behind the work in this thesis is the fact that by applying modern time-frequency techniques to the analysis of heart sounds it is possible to take advantage of the improved electronic collection of sounds while maintaining a clear relation to auscultation techniques. By maintaining this connection it is envisaged that the advantages of electronic data capture and time-frequency diagnosis can be combined to produce a new improved diagnostic aid.

1.1.2 The uncertainty of origin

Although there has been, over the years, an improvement in the understanding of the operation of the heart and causes of dysfunction, there is still a distinct lack of definite knowledge of the origin of sounds. A strong connection has been found between various pathological conditions and the modification of heart sounds, hence, the success of auscultation as a diagnostic aid. These clear empirical results are in definite contrast to the almost complete lack of a definite theory to the origin of the sounds.

In all systems that produce output signals, an understanding of the signal itself can be gained from a knowledge of the mechanism generating it. Often, this information is gained from dismantling the system or inspecting its internal structure during operation. In the case of heart sounds, the generating mechanism (the heart) provides distinct practical problems against such an inspection, not least of which is its inaccessibility. Some researchers have attempted to overcome this problem via open chest analysis of canine subjects [4][5]. The sounds recorded in these surveys are very clear but results suffer from two major limitations. Firstly, relatively low population sizes studied, due to the impracticalities involved in data acquisition, mean results are inconclusive. Secondly, the results presented must be considered in
light of the significant disturbance to the chest-thorax system required to facilitate direct access to the heart.

A second motivation behind the work in this thesis has been the potential that modern time-frequency analysis techniques have for the investigation of sound origin. It is envisaged that the application of high resolution, high sensitivity analysis techniques to an extensive and varied set of patient types will provide evidence of the sound modulation caused by varying valve type, construction and condition. With an understanding of these effects it is quite possible that evidence supporting, and disproving various hypotheses for the origin of sounds will become apparent.

1.1.3 Heart valve degeneration

There is a small but significant percentage of the population that will, at some point in their lives, have problems with the operation of their cardiac valves. In some situations, such as infant rheumatic fever, the condition is, to a degree, predictable. However, in many people valvular problems first appear as simple symptomatic features such as an unexpected shortness of breath. After the onset of these symptoms, patient prognosis rapidly decreases with time. Hence, it is important, at an early stage, to produce a clear analysis and definition of the condition of the patient’s cardiac valves. To this aim, the physician uses a number of tools to provide him/her with information on which to make an accurate diagnosis. At present, the tools available to the physician are:

1. **Auscultation**: a diagnosis based on an analysis of the heart sounds, murmurs and other pathological sounds using a stethoscope.

2. **Electrocardiogram (ECG) Analysis**: recording of the electrical activity of the heart via the connection of electrodes to the patient’s limbs and chest.

3. **Echocardiography**: a two or three-dimensional time-elapsed picture of the internal mechanisms of the thorax is generated from transmitted, reflected and received sonic waves.

4. **Cardiac Catheterisation**: with the patient under a local anaesthetic a flexible tube is inserted into the circulatory system and fed towards the valve of interest. Catheter instrumentation allows the collection of various blood flow and pressure information in the vicinity of the dysfunction.

5. **Angiography**: a radiopaque substance is introduced into the patient’s blood stream and a radiographic analysis performed. The resultant picture of the circulatory system can
be used to indicate the integrity of valves or other circulation components.

6. **Cinefluoroscopy:** a technique similar to angiography where time-elapsed pictures of the circulation of a radiopaque substance are produced.

All these tools have very distinct uses and, as such, exhibit various advantages and disadvantages. Importantly, they all provide information from which the physician can make a decision as to the nature of a particular patient’s condition. Possibly the most important of these methods is auscultation as it is used as the first indicator of the patient’s condition and methods 2 - 6 are only employed after dysfunction is suspected. In this respect, it can be seen that an improvement in auscultation, more than any other technique, will result in improved early detection of valve dysfunction.

In cases where valve dysfunction is severe it is often necessary to replace the valve with a mechanical or biological prosthesis. These operations are obviously problematic and the subjects receiving new valves are monitored continuously up to 21 days after surgery and at regular intervals thereafter for the remainder of their lives. The limited durability of bioprosthetic valves [6] and fallibility of metallic prostheses [7] is well documented. The post-implantation evaluation of the prosthesis is achieved again using the tools listed above. Again, auscultation has distinct advantages over the other techniques. Specifically, auscultation does not require expensive equipment or extra specialised personnel; it is simple and non-invasive. The definite advantages of auscultation over other techniques is again a motivation for the investigation of the possibility of time-frequency techniques as an improvement upon auscultation diagnostic procedures.

**1.2 Biomedical Signal Processing**

The processing of a signal can be motivated by a number of requirements. In biomedical signal processing the most common recurring requirement is one of extracting important information from surrounding unwanted signal components. With a complete knowledge of the wanted and unwanted signal components this improvement in signal-to-noise ratio can often be achieved with careful choice of recording procedure and equipment. In the case of heart sounds, contamination of the desirable signal can often occur prior to recording. Recordings of sounds are made externally, at a distance from the heart on the rib cage that quite probably acts as a resonant chamber for a whole range of other noisy physiological phenomena.
Chapter 1: Introduction

Probably the most significant limiting factor in the analysis of heart sound signals is the variation of every aspect of this signal across common subject classes and often, within the individual. It is almost impossible to account for the vast majority of personal and situation based factors that can have, to varying degrees, an effect on the human physiological operation. To this end, biomedical signal processing presents a number of unique problems with respect to the collection of data, data processing and the evaluation of results.

1.3 Contents of the Thesis

This thesis describes the work performed by the author in the investigation of the time-frequency analysis of native and prosthetic heart sounds.

Chapter 2 provides an introduction to the signal and analysis methods used in this work. The heart sound signal is introduced as a consequence of the action of the heart and the circulation of blood. A number of historical viewpoints on the structure of the cardiac sound cycle and the origin of sounds are presented. Consideration is also given to the case where the heart is functioning abnormally and hence, pointers to diagnostic applications are discussed. This chapter concludes with a comprehensive review of past work performed in the field of automated and computer-based heart sound analysis.

Chapter 3 describes the elements of the hardware system developed for the acquisition of data used in this study. The chapter commences with a description of the overall system, followed by details of the purpose built pre-processing circuit and the phonocardiographic transducer. Particular attention is paid to the definition and justification of the pre-processing and transducer frequency bandwidth. The second half of the chapter moves away from the hardware and instead concentrates on describing the logistics involved in data acquisition. A description of the procedures employed during recording is given together with details of the distribution of populations analysed in the study. Finally, the chapter concludes with the illustration of a number of typical recordings taken from subjects in the various population groups.

Chapter 4 describes the theory of time-frequency signal analysis. The four techniques considered are: the short-time Fourier transform (STFT), the wavelet transform (WT), the Wigner distribution (WD) and the Choi-Williams distribution (CWD). After describing the theory behind these transforms, a practical analysis of their performance is presented.
Chapter 1: Introduction

Relative performance of the four techniques is investigated via their application to two simple sinusoid test signals and to a modelled heart sound signal.

Having made a decision in chapter 4 as to the most suitable transform for heart sound analysis, chapter 5 describes the results obtained for this technique's application to the recorded sounds. Chapter 5 includes typical time-frequency results generated from the analysis of data from all population types. Alongside these typical results, a statistical analysis of the results for the full population for each valve type is presented.

The discussion in chapter 5 provides an analytical analysis of the data. Chapter 6 describes the methods and results for an alternative automated analysis. The chapter commences with a description of the techniques used to perform feature extraction from the time-frequency data. Following the theory of feature extraction, a description of the classification techniques used for the identification of various population groups is presented. Results are given for classification success of the techniques applied to the identification of native valve condition, native valve origin of dysfunction and bioprosthetic valve condition.

Chapter 7 provides a conclusion to the results presented in the previous chapters. Particular attention is paid to the analysis of results with respect to the initial aims presented in chapter 1. Finally, suggestions are made for possible areas of further research.
Chapter 2

Introduction to Heart Sound Signal Analysis

The analysis and transformation of a signal is, in general, performed to facilitate the extraction of information. This is particularly the case in biomedical signal processing. This chapter provides an introduction to the concepts investigated and employed in this thesis for extracting information from surface-recorded heart sounds. Initially, a description of the signal of interest (heart sounds) and their generating mechanisms is given. After describing the signal to be analysed, a description of the basic concepts of time-frequency analysis is presented. The chapter concludes with a comprehensive review of past work in the field of automated and computer-based heart sound analysis.

2.1 Heart Sounds

Prior to the analysis of a signal it is often desirable to gain an understanding of the mechanisms that influence the generation of the signal. As in many signal processing applications, a great deal of information about the composition and behaviour of a signal can be divulged from an understanding of the mechanisms producing it. In this specific case the sound-generating mechanisms are blood flow and the action of the heart. The following sections provide an introduction to the operation of the heart, as well as a number of historical viewpoints on the origins of sounds.
2.1.1 Circulation

Figure 2.1 shows an illustration of the layout of the cardiac circulatory system. A description of the components of this system is given in [8].

![Diagram of the cardiac circulatory system](image)

**Figure 2.1:** The circulation system.
Basically, the heart pumps blood into the arteries, which subsequently split into capillaries. From the capillaries blood is collected up into veins which return to the heart. The pumping of blood via the heart is achieved in four basic stages which are described below and illustrated in Figure 2.2.

(a) The left and right atrium fill with blood returning from the lungs and body respectively.

(b) Rising pressure in the atrium, due to returning blood, forces open both atrio-ventricular valves (tricuspid valve in the right side of the heart and the mitral valve in the left side of the heart). As a result of these valves opening, blood flows into the ventricles.

(c) When both the atrium and ventricles are full of blood the heart begins to contract. Contractions originate at the right atrium and after a short period spread to the left atrium. Contraction of the atrium forces any remaining blood into the ventricles.

(d) Contraction spreads from the atrium to the ventricles, aortic and pulmonary valves open and blood is ejected from the heart.

![Figure 2.2: The four stages of the heart's operation: (a),(b) diastole; (c),(d) systole.](image-url)
Chapter 2 : Introduction to Heart Sound Signal Analysis

Figure 2.2 illustrates a number of possible origins to heart sounds, including the movement of blood in vessels, chambers, and through valves. Other possible origins of sounds are the rapid contraction and relaxation of the heart chambers together with the subsequent transition of valves. The exact origin of sounds whether due to valvular transitions or blood column vibrations is debatable [9][10][11].

2.1.2 Sound origins

A proposal for the origins of heart sounds was given by Luisada in 1972 [12]. The limitations of the transducer had been overcome by the use of phonocardiography techniques and an attempt to monitor the heart's operation was achieved by measuring intracardiac pressure. It should be noted that a great deal of Luisada's research was substantiated using dogs as subjects. The basic structure of the heart is common to all mammals and the dog's heart will exhibit the same sound components as a human. Figure 2.3 shows an illustration of the heart sounds proposed by Luisada.

Figure 2.3: The Luisada sounds.

These "Luisada sounds" contain four basic components (I - IV) and in fact Luisada also tentatively postulated the occurrence of a fifth. It is expected that in a normal adult only the first and second sounds will be audible. The other sounds are often accentuated in various age groups and pathological conditions. The sub-division of the sounds in Figure 2.3 is based on both empirical and theoretical ideas. As already noted, during normal operation the appearance of all of these components is not expected and often, in low intensity situations, components merge into one.
The descriptions of sounds given by Luisada are all based on the assumption that sounds are not due to valve movements. His evidence for this assumption was based on experimental results showing a pressure change across valves (assumed as valve closure) occurring 30-35 ms before the associated sound [13]. It has since been shown that although a pressure change occurs the valve will not close instantaneously and timing may still be maintained [14].

At the same time as Luisada was attempting to explain the origin of heart sounds, alternative ideas were being cultivated in the U.K. by Leatham and Leech [15] that maintained the argument that the major heart sounds originated from the action of the heart valves. The original results of this work were presented in the early 1950's [16]. The theories presented were validated by Leatham and Leech at a later date using modern echocardiography technology [15] and by Laniado et al. using x-ray techniques to photograph cusp movement [14]. Figure 2.4 shows an illustration of the sounds described by Leatham.

![Illustration of Leatham sounds](image)

Figure 2.4: The Leatham sounds.

It can be seen that the four basic sounds are the same as Luisada's sounds (Figure 2.3) but the interpretation of their structure and origin has changed to accommodate the alternative results. Also, the various components of the three basic sounds differ quite considerably. This difference in component structure may be accredited to the fact that heart sounds are of very short duration, low intensity and vary extensively from subject to subject. Orias [10] commented in 1949 on the results presented by past investigators; his comments are equally relevant to results presented since.

"The first sound is a complex phenomenon which causes most of the discrepancies existing between findings of different investigators".
"Each investigator using some special technique has taken account only of the factors that his own particular method revealed and has ignored the existence of all that it was unable to detect".

Another possible explanation for the discrepancies between researchers' findings is given by Dayem and Raftery [17].

"The number of phonocardiographic components discernible in a tracing of the first heart sound vary with the characteristics of the filter used and the enthusiasm of the interpreter".

It was originally suggested that the first heart sound exhibited two major components, originating from the tricuspid and mitral valves [18]. As can be seen from Figure 2.1, the left side of the heart supplies blood to a much larger circuit than the right and hence, it is reasonable to assume that the mitral left side is the dominant sound generator. In fact, it has been demonstrated that the origin of the first sound is predominantly, if not totally, left-sided [19]. Various factors have been proposed as mechanisms that effect the integrity of the first heart sound.

- Adequacy of the atroventricular cusps to halt regurgitant flow.
- Mobility of cusps.
- Position of cusps at the onset of systole [20][21] (the position of cusps will determine the closure delay and hence the amount of ventricular pressure on closure).
- Rate of ventricular contraction [20] (assuming ventricular pressure to be the source of sound energy, higher pressures will cause increased cardiovascular vibration).

The justification of these four factors is, in the majority, based on their exaggeration or suppression in various pathological conditions. Variation of first sound intensity with leaflet closing velocity was investigated by Hearn et. al. [22]. It was found that sound intensity was dependent upon mitral valve closing velocity only within a specific frequency range. After replacement of the mitral valve Dayem and Raftery [17] observed that the high frequency component of the first sound had been replaced by prosthetic valve components. From these results it was postulated that the mitral valve contributes to a large percentage of the first sound.

The end of systole is marked by the closure of the aortic and pulmonary valves. This dual action is assumed to cause the two components of the second heart sound. The semilunar valve origin of this sound was proposed well before the phonocardiographic era, but is now supported by experimental observations. The characteristics and pathological significance of
the sound and its components reported by Leatham in 1954 [16] are very close to the modern interpretation. An attempt to describe the factors that affect the intensity of the second sound is given in [23]. The work supports previous ideas for the hemodynamic factors determining the intensity of the second sound and introduces a number of possible anatomical factors. The proposed factors governing second sound integrity are:

- Rate of change of diastolic pressure gradient that develops across closed aortic and pulmonary valves (blood pressure relates directly to the forces exerted on the valve cusps).
- Distensibility of cusps (flexibility).
- Valve weight (lighter valves will close faster).
- Valve surface area (a large sound radiating surface).

In addition to the major first and second heart sounds there is definite evidence of low intensity third and fourth cardiac heart sound. It is generally agreed that the low-pitched diastolic third sound has a non-valvular origin. Luisada [12] postulates that the third sound is due to the transition of the ventricles from active to passive distension. Leatham and Leech [15] accredit the sound to the sudden halting of inflowing blood as the ventricles become full. The appearance of the third sound in young adults, especially athletes, is not unusual but in older patients the appearance of a significant third sound is often an indication of cardiac dysfunction. Again, the fourth heart sound is thought to have a non-valvular origin. Leatham and Leech [15] state that the sound is too late to have an atrial origin and may be caused by ventricle filling before contraction. The fourth sound, a very quiet sound, may be present in normality but, as in the case of the third sound, it becomes accentuated in specific pathological conditions. Such conditions include ischaemic heart disease and systemic hypertension [24]. In conditions of tachycardia the third and fourth sounds may both be present but due to their proximity they are often inseparable.

The four sounds described above are not the only sounds that can be found in the cardiac cycle. As detailed by Leatham (Figure 2.4) in addition to sounds associated with valve closure there are also sounds associated with opening of the cardiac valves. The ejection sound and the opening snap are coincidental with opening of the semilunar valves and atrioventricular valves respectively. Both these sounds, although commonly accredited to valve cusp movement, have also been associated with blood flow through a diseased valve [24].
2.1.3 Discussion

The proposals for the origin of sounds given by Luisada [12] were based on the idea that sounds and valvular events are not coincidental. Although this assumption has been proved inaccurate, Luisada's work illustrates the important concept that the heart must be considered as a whole vibrating object. This point is also highlighted by Waider and Craige [18]. The coincidence of valve closure and the first heart sound indicates a temporal relation but whether valve vibration [14] or blood and tissue vibrations [25] cause the sounds is debatable. It has been demonstrated that valves and attached chordae alone cannot account for the total surface sound intensity. If intraventricular blood mass and its boundary structures (valves and ventricular muscle) are included in the calculation, MacCanon et. al. [25] claim that sound production is feasible. The transmission of sound along arteries was investigated by Farber and Purvis [26]. It was found that the principle method of vibration transmission was due to transversal vibration of arterial walls; the physical connection of these walls to the heart structure suggests the possibility of system vibration. In spite of a great deal of research directed at uncovering the true origins of sounds, the description given by Orias in 1949 [10] is very close to the modern interpretation [24][11]. A point that Orias did get wrong was that future advances would improve our understanding but, in many cases, this has simply added to the debate.

2.2 The Dysfunctioning Case

To this point the descriptions of sounds and their generation has concentrated on the normal case. As one of the objectives of this work is to use time-frequency techniques as a tool for description and classification of normal and abnormal sounds, it is important that we describe both these cases. In cardiology there are a great number of situations in which the heart deviates from its normal operation. These include disruption of the heart's electrical system, damage to myocardial muscle performance, disturbance to the heart's blood supply and degeneration of heart valves. In this work we have concentrated on the final case referred to in medical texts as valvular heart disease.
2.2.1 Valvular heart disease

Any of the heart's four valves may become diseased but in the vast majority of cases it is the mitral or aortic valves that become defective. Hence, this study, for reasons of application and logistics, concentrates on the analysis of subjects with defects in these two valves.

The causes of valvular heart disease are numerous. An introduction to the major factors is presented in [27] and more detailed descriptions are presented in the medical texts [28][29]. The most common effects of valvular disease are stenosis and regurgitation with the first condition often preceding and, perhaps, causing the second. The most common audible effect of these two conditions is the appearance of a murmur in the cardiac cycle. Murmurs can be produced by both valve stenosis and regurgitation. Stenosis can usually be credited as the cause of a murmur during the valve's open period. Conversely, regurgitation is the result of a valve leaking during its closed period. Considering Figure 2.4 and remembering that the ejection sound and opening snap coincide with aortic and mitral valve opening respectively, the origin of diastolic and systolic murmurs can be associated with specific aortic and mitral conditions. During systole the mitral valve is closed and the aortic valve is open, hence, a systolic murmur may be due to either mitral regurgitation or aortic stenosis. During diastole the mitral valve is open and the aortic valve is closed hence, a diastolic murmur may be due to either mitral stenosis or aortic regurgitation.

Figure 2.5 illustrates the natural history of the condition aortic stenosis, as presented by Goodfield and Bloomfield [27]. The figure shows that during initial development of the condition there is only a small increase in patient fatality as compared to the normal population but after onset of symptoms there is a catastrophic drop in patient survival. The graph illustrates the necessity of early detection of such conditions at the onset of symptoms or possibly, via a screening programme, before obvious symptoms appear.
2.2.2 Treatment of valvular heart disease

With the early detection of valve dysfunction, a course of drugs can be used to alleviate symptoms and increase the natural life of the valve. In severe conditions, shown on the right-hand side of the graph in Figure 2.5, prognosis is low and valve surgery is often a necessity. The treatment of dysfunctioning valves takes one of two forms: valve repair or valve replacement. A replacement valve can be one of two types: a mechanical or biological prosthesis. Most common of the mechanical prostheses are ball-and-cage valves such as the Starr-Edwards and tilting disc valves such as the Bjork-Shiley. Bioprosthetic valves, such as the Carpentier-Edwards (C-E) valve, are made from pig valves mounted on a stainless steel frame. Figures 2.6a and 2.6b show photographs of a Bjork-Shiley mechanical prosthesis and a C-E bioprosthetic valve respectively.

Figure 2.5: Typical survival of the normal population and patients with aortic stenosis.
Chapter 2: Introduction to Heart Sound Signal Analysis

Figure 2.6: Prosthetic heart valves: (a) Bjork-Shiley tilting disc mechanical prosthesis; (b) C-E porcine bioprosthesis.

An important factor to be considered is the durability of these replacement valves. A study into the life span of a number of valves including biological and mechanical examples was performed over a period of twelve years at the Royal Infirmary of Edinburgh (RIE) [6]. Results of this study indicated that survival of patients with a Bjork-Shiley valve dropped linearly with an increase in time at about 2.6% per year. In contrast, the porcine valves included in the study showed an ever increasing fall in the rate of survival as time increased. After thirteen years of study the survival rate for porcine-valve subjects was approximately 16% less than the equivalent mechanical valve subjects. The results suggest that complications such as thrombosis effect both mechanical and biological prosthesis, forcing an underlying linear drop in valve duration. The increased incidence of valve failure in the porcine case is quite probably due to the degeneration of the biological tissue used in bioprosthetic valve construction. It is valid to say that implanted bioprosthetic valves, because of their tissue...
construction, begin gradual degeneration from the moment they are implanted. Due to the obvious robust construction of the mechanical prosthesis similar degeneration does not occur, resulting in the superior life span recorded in the Edinburgh study [6].

For similar reasons to those given in the case of degenerative native valves, the development of techniques and equipment that provide an aid to early detection and diagnosis of dysfunctioning prosthetic valves is highly desirable. Later in the thesis results are presented for the analysis of normal mechanical valve subjects. A comparison of normal and dysfunctioning mechanical valve cases was impossible due to the scarcity of subjects suitable for recording with diagnosed abnormal mechanical valves. The reasons for this lack of suitable subjects were two-fold. First, as already described, the survival rates of mechanical valves are higher than bioprosthetic valves, hence a large population of degenerated mechanical valves does not exist. Secondly, the failure of mechanical valves is often catastrophic resulting in patient mortality [30][31], again resulting in a reduction in the available patients suitable for recording. Such catastrophic failure in the case of the Bjork-Shiley convex-concave valve has gained much media coverage due to the phantom nature of failures [7]. Reoccurring strut failure found in implanted Bjork-Shiley convex-concave valves has never been successfully reproduced experimentally. There are a great number of people with this valve implanted which points to the possible application of the techniques in this work to the monitoring of this high risk group for initial signs of valve failure.

2.3 Introduction to Time-Frequency Signal Description

The description of a signal as a linear weighted sum of harmonically related sinusoids was first presented by Fourier in the early 1800's. Since this definition of signal spectral analysis, the Fourier transform and its more practical successor the fast Fourier transform (FFT) have been used extensively as an tool for the description of signal spectral characteristics. It is clear from this historic success that there is definite value in the description of a signal in terms of its frequency spectrum and conversely, in terms of its time-domain description. However, a new perspective on signal description was presented by Gabor in 1946 [3]. Gabor postulated that the time-domain and frequency-domain forms of a signal were simply extreme examples of a more general time-frequency description. Gabor also expressed the opinion that a compromise between these two extremes was of more value due to its ability to extract signal information with respect to both time and frequency. Although Gabor's work has been described as a new perspective on the analysis of signals, one particular form
of time-frequency signal description has been used for centuries. Figure 2.7 shows the first two bars of the piano solo in Grieg’s piano concerto.

![Figure 2.7: First two bars of the piano solo of Grieg's piano concerto.](image)

The time-frequency description shown in Figure 2.7 may be described as a discrete time, discrete frequency representation. As is well known the evolution of the signal (music) is from left to right and the frequency (pitch) of the signal is represented by the vertical position of notes on the stave. This format closely matches the format of the modern time-frequency representations described and illustrated in this thesis. Figure 2.8 shows the time-frequency plot of a signal containing four consecutive notes generated using STFT.

![Figure 2.8: STFT of a signal composed of 20 Hz, 40 Hz, 30 Hz, and 10 Hz sequential components.](image)
By comparing Figures 2.7 and 2.8 it can be seen that the STFT is displayed in a very similar format to the two bars in Figure 2.7 but provides a continuous description in both time and frequency. If the power spectral density were to be generated for the piece shown in Figure 2.7 it would be valid to comment that such a description provides all the correct notes but they will have been reordered. Hence, although providing information about the signal, the power spectral density is not the whole story and is by no means Grieg’s piano concerto. The suitability of time-frequency as a signal descriptor, as pointed out by Gabor, is emphasised by our everyday experiences. The time-frequency description of the piece in Figure 2.7 is a necessity because of the auditory system’s perception of sound in terms of time and frequency. This thesis is directed at the analysis of heart sounds and the present method used to evaluate heart sounds is auscultation. The use of an auditory analysis of heart sounds renders the resultant description inherently time-frequency. The universal success of auscultation is a possible indicator of the suitability and potential of time-frequency methods for the description of recorded heart sounds.

2.4 Time-Frequency and Spectral Estimation of Heart Sounds: Past Work

Gabor’s original time-frequency ideas have in the past been applied to the analysis of heart sounds. The 1954 article by Geckeler et. al. [32] describes the application of Bell Telephone Laboratories’ equipment for the plotting of spectrograms of sounds recorded from over 100 subjects. The aim of the work was to investigate and define spectrogram normality in relation to auscultation definitions. This early pioneering work made a number of astute observations on the nature of the problem and the possibility of a solution. Firstly, the complexity of the problem was identified; in particular, it was noted that the idea of normality was not a term that could easily be defined. Examples used in the study that were classed as "normal" showed large mutual variation and often variation from accepted auscultation definitions of normality. A significant conclusion made by this work is the need for "clearer and cleaner" patterns. This need can be more precisely defined as a need for higher resolution. The results illustrated in [32] show little, if any, localisation in frequency. The calculation of time and frequency resolution in the work by Geckeler et. al. [32] would require a precise definition of the equipment and techniques used but results presented suggest a frequency resolution in the range of 100 Hz and a time resolution of about 50 ms. Quite probably because of these significant resolution limitations Geckeler et. al. conclude that their cardiospectrograms were not a viable alternative to auscultation, but could be applied as a teaching aid or prove
Chapter 2: Introduction to Heart Sound Signal Analysis

valuable in screening programs. Little if any advancement on the results presented by Geckeler et. al. was made in the following two decades. Winer et. al. [33] presented a "new method" (the contour spectrogram) for the analysis of heart sounds. An improvement in clarity of results was achieved but due to the common technological base, no improvement in resolution was achieved over the results presented by Geckeler et. al.. The contour spectrograph was still finding application to heart sound analysis as late as 1969. In the paper by Hylen et. al. [34] the contour plot is applied to the diagnosis of aortic ball variance in patients with implanted aortic Starr-Edwards prosthesis. This work shows a high level of success for the identification of subject ball variance. The significant limitation of the study, as admitted by the last author, is the lack of resolution in the technique. Also of note is the fact that the features used in identifying ball variance were exclusively frequency features. Time information in the results is only passively used to identify the significant features in the cardiac cycle.

Although time-frequency methods had been identified as the most promising and intuitive method for non-stationary heart sound analysis, their lack of resolution was a major problem. A solution to this resolution problem was not forthcoming and research turned to the analysis of sounds in the frequency-domain only. In the early 1970's Kingsley [35] and Gordon et. al. [36] used analogue filter techniques to analyse the spectral content of both metallic and biological prosthesis. The techniques used in these two studies were a one-dimensional equivalent of that used by Geckeler et. al. [32], Winer et. al. [33] and Hylen et. al. [34]. These very poor resolution techniques were used to analyse heart sounds right up to 1979 [37][38]. The improvement in digital computers and their increased availability resulted in the application of digital techniques to the problem of heart sound analysis. A number of researchers applied the FFT to a varied array of heart sound applications [39][40][41][42][43][4]. During this period one or two more attempts were made to analyse recorded heart sounds on a time-frequency basis. In 1977 Kagawa et. al. [44] reverted back to the analogue contour sound spectrogram to analyse 50 patients with diagnosed dysfunctioning prosthetic valves. Results were inconclusive due to the lack of a normal control set and, as already discussed, the inadequacy of the analogue spectrogram technique. A novel attempt at time-frequency sound analysis was presented by Iwata et. al. [45]. In this work Iwata et. al. used an autoregressive (AR) parametric technique to analyse the spectrum of successive sections of the recorded sound. The AR method, unlike the analogue spectrogram and FFT, is not limited in resolution by the time window over which the signal is being analysed [46]. The disadvantage of these AR methods is that model order must be carefully estimated to prevent estimation inaccuracies. In the work by Iwata et. al. [45] a constant model order is employed and no
consideration is given to the possible variation of model order as different signal segments are analysed. Recent work on the application of parametric spectral estimation techniques to heart sound analysis has demonstrated the importance and difficulty of estimating the true model order for heart sound analysis [47][48][49]. Iwata et. al. continued their work to include a scheme for automatic segmentation of the recorded sounds [50] and finally combined their results to produce a completely automated heart sound classification scheme [51][52]. To the author's knowledge these are the only presented results for the development of such a comprehensive automated diagnostic tool. During the 1980's and to this day the FFT remains the most popular technique for heart sound analysis. Papers published that exclusively used the FFT as an analysis tool include [53][54][55][56][57][58][59][60]. Alongside these works exclusively using the FFT, a number of researchers investigated the usefulness of advanced parametric spectral estimation techniques. As already mentioned Iwata et. al. [51][52] used an AR parametric analysis technique. Four years after the publication of these results Nandagopal et. al. [61] published work that compared Iwata’s AR technique to the FFT and demonstrated its higher resolution when applied to normal heart sounds recorded from 17 children. Although Nandagopal et. al. demonstrated a higher resolution for Iwata’s AR technique again, no consideration was given to model order selection. An alternative deterministic parametric technique that has proved useful in the analysis of heart sounds is Prony’s method. In the late 1980’s Koymen et. al. used Prony’s method to analyse mechanical valve sounds in normal subjects [62][63][64][65]. More recently a modified version of Prony’s method has been demonstrated to be superior to FFT techniques and has been successfully applied to the classification of normal and abnormal valves [48]. Another parametric technique previously applied is the Steiglitz-McBride method, first used by Joo et. al. in 1983 [66] and updated in 1984 by BeMent and Veeneman [67]. A comparison of this technique to the FFT [68][69] showed that, depending upon the feature of interest, both methods could outperform the other. This work by Cloutier et. al. [68][69] and later that by Durand et. al. [70] used resultant classification success as the measure of the performance of the spectral estimation technique. This is in contrast to the work by Foale et. al. [71] who applied the FFT and a AR parametric technique to sounds and concluded the AR technique to be superior due to its better resolution. In the most recent publications by Durand et. al. in this field [58][70] it has been shown that the FFT with a rectangular window produced the best classification results for a normal/abnormal test set. Hence, recent work by these researchers has exclusively used the FFT for spectral estimation and has moved towards comparing classification techniques rather than spectral estimation techniques [58][72].
In the early 1990's two papers appeared that used the STFT for the time-frequency analysis of heart sounds. Picard et. al. [73] used the STFT to analyse the sound produced by various mechanical valve types while Jamous et. al. [74] concentrated on the optimisation of the STFT technique to heart sounds.

Although the STFT technique was an improvement upon the analogue spectrograms presented in the 1950's and 1960's, simultaneous high time and frequency resolution remains the "holy grail" of the subject. In 1992 an exciting new prospect for the analysis of heart sounds was presented by Wood et. al. [5]. It was shown in this work that using the newly developed Cohen class of time-frequency distributions a clearer understanding of the origin of the first heart sound could be divulged. Although the work presented by Wood et. al. was not a motivation for the work in this thesis, it did provide clear evidence of the suitability of Cohen's class of time-frequency distributions to the analysis of heart sounds. Since the commencement of the work in this thesis, a number of other researchers have begun work on the analysis of heart sounds using modern time-frequency techniques [75][76]. These investigations have concentrated on the analysis of the suitability of various advanced time-frequency techniques to heart sounds analysis. To the author's knowledge, this thesis is the only published work that demonstrates diagnostic results and comments on the use of advanced time-frequency techniques for the analysis of a significant group of cardiac patients.

The above description of the past work in this area of research has concentrated on the evolution of signal processing techniques that have been adopted by various researchers. Another important area of the work that has evolved alongside the signal processing is the specific applications. In the early works [34][35][36][44] there seemed a common goal of identifying dysfuncting metallic prosthesis. This was possibly due to the clear understanding that these relatively new metallic valves generated high frequency click sounds obviously associated with valve operation. Hence, monitoring valve integrity via sound analysis was seen to be a distinct possibility. In the majority of cases techniques were applied to the analysis of subjects with implanted ball-and-cage valves in an attempt to identify the abnormality known as ball variance. Ball variance can be defined as a valvular dysfunction resulting from changes to the ball shape integrity and even size. As valve design has improved over the years it has been demonstrated that, in the majority of cases valves, based on a tilting disc principle provide better hemodynamic properties than original ball-and-cage valves. Hence, cases of ball variance are few and the need for techniques to identify ball variance has almost disappeared. The modern popularity of mechanical disc valves has resulted in a number of researchers attempting to analyse their sound signatures. In-vivo studies of these valve types
have been performed by Durand et. al. [60] and Koymen et. al. [64]. In the work by Koymen et. al. [64] and a number of similar studies, the analysis of valves is only performed for a normal population and not a full normal/abnormal set. In Durand's work the analysis of normal and abnormal valves is made but it should be noted that the analysed valve sounds were recorded from sheep who had been implanted with normal and pre-doctored abnormal prosthesis. Studies investigating the spectral analysis of modern disc valves suffer from the lack of availability of subjects suitable for recording with implanted abnormal valves. As described in section 2.2.2 the abnormal mechanical valve population is low because of extended valve durability and often catastrophic failure. A number of other researchers have tried to overcome these logistic problems by testing valves in an in-vitro situation [57][59]. Both these investigations concentrated on the analysis of the Bjork-Shiley convex-concave valve because of the specific problems it exhibits [7].

Another area in which mechanical valve sounds have been analysed is with respect to the noise level of associated clicks. Despite the improved durability characteristics exhibited by mechanical prosthesis, it has been postulated that the detrimental psychological effect that a continuously clicking valve can have on a patient renders silent bioprosthetic valves a more desirable treatment [77]. For this reason a number of researchers have investigated the noise level generated by metallic valves and have concluded that the design of new prosthesis should take into consideration emitted sound intensity [53][73][78].

In addition to the work described above, the analysis of heart sounds has been extensively applied to the task of monitoring subjects with implanted bioprosthetic valves [42][43][79][66][58]. The justification of these pieces of work is that bioprosthetic valves, due to their tissue construction, degenerate from the day they are implanted. This continuous state of degeneration results in the inevitability of valve failure between ten and fifteen years after implantation [77]. The limited durability of bioprosthetic valves results in a definite need for accurate post-operative monitoring. Hence, the search for non-invasive evaluation techniques. There are a very large number of patients at present with implanted bioprosthetic valves who undergo regular evaluation sessions in an attempt to achieve early detection of valve degeneration. This situation provides a definite need for techniques that can be used to monitor bioprosthetic valve integrity. Despite this present need, it is envisaged that as the metallic valve prosthesis becomes more popular the need for bioprosthetic valve analysis techniques will decline.

25
The spectral analysis of both mechanical and biological prosthesis, as described above, has been extensively investigated. In contrast, there has been very little work performed on the analysis of native heart valve sounds. A few studies have been performed on the analysis of normal native valve populations [39][40][55][61] but the comparative analysis of normal and abnormal native valves has been almost completely ignored. At present auscultation is used extensively by medical practitioners, as an aid to the evaluation of native valve condition. Hence, there is a distinct possibility that time-frequency signal analysis may be useful in the identification of native heart valve condition.

### 2.5 Summary

The first two sections of this chapter were intended to provide the reader with an introduction to the signal of interest and the mechanisms governing its production. From the discussions in section 2.1 it can be seen that in the past there have been various hypotheses presented to explain the origins of heart sounds. Despite these investigations there is still a definite lack of knowledge as to the true origin of sounds. In the analysis of time-frequency as a technique for the evaluation of valve integrity it is important not only to understand the operation of the normal heart but the modification of its operation during dysfunction. Section 2.2 provides a description of the various cases of valve dysfunction and their treatment. During this discussion it was identified that there is a need for techniques that can provide information concerning valve integrity. The specific applications identified were early detection of native valve abnormalities, monitoring of implanted bioprosthetic valves and the monitoring of problematic metallic valves, in particular the Bjork-Shiley convex-concave prosthesis.

Section 2.3 provided a very simple introduction to the concepts of time-frequency signal description and analysis. Using an illustrative example the concepts of time-frequency with relation to more traditional time only and frequency only descriptions was presented. It was also suggested that due to the inherent time-frequency format of auscultation, the application of time-frequency signal analysis techniques to recorded heart sounds offers potential as a diagnostic aid.

Finally, a detailed description of past work in the field of heart sound spectral estimation, both time-frequency and frequency only, was presented. From the discussion in section 2.4 it can be seen that there are areas that have yet to be fully investigated. In terms of analysis techniques, it was found that the complete analysis of time-frequency as an aid to the diagnosis of a significant valve population has yet to be investigated. Also, there has been no past
investigations into the analysis of a complete native valve normal/abnormal population using any spectral estimation technique.
Chapter 3

Data Acquisition

This chapter describes the hardware, software and recording procedures used for the acquisition of heart sounds from various subjects. Particular attention is paid to the description of the frequency bandwidth of the recording hardware, the purpose built pre-processing hardware and the phonocardiographic transducer. Also given is a description of the recording protocol used to ensure consistent data capture and reproducibility of results. Finally, a description of the various subject populations is followed by a number of examples of typical PCG and ECG recordings.

3.1 Data Acquisition Hardware

To facilitate the collection of heart sound data an application specific hardware and software recording system was developed. Figure 3.1 shows a block diagram of the overall data acquisition hardware system.
Figure 3.1: Overall structure of the hardware system used in heart sound data acquisition.
The recording of patients was performed at one of four sites: the wards of the RIE, the cardiology unit at the RIE, the Edinburgh Astley Ainslie Hospital and at the University of Edinburgh, Department of Electrical Engineering. The recording of data at these varied sites required the equipment to be portable hence, the choice of a portable laptop computer. The system shown in Figure 3.1 has three main components: the portable computer / analogue to digital converter (ADC) card, the pre-processing hardware and the cardiac transducer. The following sections describe the specification, and in the case of the pre-processing hardware, the design considerations for these three blocks. It should be noted that no description of the ECG hardware is given because this study has concentrated on the analysis of the PCG signal. It is adequate to say that ECG acquisition and pre-processing was performed using a Siemens-Elema Mingocard 3 ECG machine which provided the required clear lead II signals.

### 3.1.1 Portable computer / ADC card

The portable computer performed two main jobs: controlling the ADC card and the provision of a graphical front end to the data acquisition system. The graphical display provides a means of initiating data acquisition and inspecting the integrity of ECG and PCG transducer connections prior to signal recording. Although the graphical display tool did provide a record facility, it was found that at high sampling rates it was impossible to both display and record data simultaneously. For this reason, the laptop-based data display tool was used primarily to inspect the integrity of recording equipment and positions before the blind acquisition of data. Figure 3.2 shows a typical example of the graphical display provided by the acquisition software.

Conversion of the captured analogue heart sounds to a digital format is necessary to allow the data to be stored on the lap-top hard disc drive. This analogue to digital conversion is achieved using a Blue Chip Technology ADC42 Rev.B hardware card. Data is converted with 12-bit precision at two alternative 5 kHz and 20 kHz sampling rates.
Figure 3.2: Typical example of the graphical display provided by the acquisition software. Where CH1 and CH2 display the ECG and PCG input signals to the ADC as shown in Figure 3.1.
3.1.2 Analogue pre-processing

The aim of this stage of the recording process is to remove unwanted input data that appears outside the bandwidth of interest. By enhancing the signal in the bandwidth over which the majority of heart sound energy appears it is possible to significantly improve the signal-to-noise ratio of the recorded data. Previous studies have shown that the majority of heart sound energy lies below 1 kHz [47][41][55]; in fact, the great majority of signal energy appears in the region of 50 Hz. This low frequency specification for the bandwidth of interest has an exception in the case where sounds are recorded from subjects with implanted mechanical prosthetic valves. Due to the nature of the materials used to construct mechanical prosthesis, valve opening and closing can result in the production of significant high frequency sound components. Hence, in some in-vitro investigations, frequencies of more than 15 kHz have been reported [53].

Due to the contrasting bandwidths of the bioprosthetic and metallic prosthetic valves, the pre-processing hardware was developed with two possible recording setups. Figure 3.3 shows a block diagram of the two channel analogue pre-processing hardware.

The ECG lead II signal was obtained from the output port of a Siemens-Elema Mingocard 3 ECG recorder. This unit was available at each of the recording venues and was used as the input to the purpose-built recording hardware.
Input analogue signals from both the high frequency and low frequency microphones were initially high-pass filtered. Filtering was performed using a third order high-pass Butterworth filter with a 50 Hz cut-off frequency. The rejection of this low frequency information is justified for four reasons. Firstly, at these low frequencies the signal is dominated by noise generated by movement and breathing of the subject. Secondly, inclusion of the dominant low frequency sounds reduces the effectiveness of the recording procedure to capture clear and significant heart sound signals. The accuracy of the recorded signal is governed by the 12-bit full range quantisation applied by the ADC. If low frequency signal components were not removed the majority of this dynamic range would be used to accommodate the low frequency non-cardiac sounds. The higher frequency heart sounds would be quantised in only a fraction of the full dynamic range. Reduction of the effective dynamic range of heart sounds reduces the accuracy of the recording and increases quantisation errors and noise. Thirdly, removal of low frequency noise allows the signal of interest to be clearly displayed during recording. The prime purpose of the recording display shown in Figure 3.2 is to facilitate the inspection of sounds prior to recording. The display of sounds is particularly useful for positioning the microphone where the heart sounds appear loudest. A similar search for the position on the chest where sounds appear loudest is performed using the stethoscope during auscultation. The analysis of heart sound amplitude and hence, microphone positioning becomes problematic if the sounds of interest are swamped by undesirable low frequency noise. Finally, in past work [37] it has been suggested that the very low frequency components of heart sounds are, in the majority, due to ventricular vibrations. It is also very likely that a large percentage of these very low frequency vibrations are generated by the resonance of the chest cavity.

For both the PCG and ECG signals, the pre-processing analogue hardware performed a pre-amplification operation. The major purpose of this pre-amplification was to maximise the input signal to the ADC. The ADC adds noise to the signal during the quantisation operation; it is commonly known that the noise introduced via this process is independent of the input signal. Maximisation of signal-to-noise ratio with respect to this quantisation noise is achieved by presenting a signal at the ADC input that has maximum dynamic range. This was achieved for both the PCG and ECG signals using a variable gain pre-amplifier. The final stage of the pre-processing of the PCG signal is a low-pass anti-aliasing filter.

As already explained, it is well known that biological and mechanical valves require different recording bandwidths. In the case of the biological valve recording hardware (channel 1), the upper cut-off frequency was set at 2 kHz. Channel 2 of the PCG pre-processing hardware facilitates the recording of sounds from subjects with implanted mechanical prosthesis. As
already mentioned, sounds generated by the operation of mechanical prosthesis can extend up to and above 15 kHz. The recording system, however, has a low-pass cut-off point at 10 kHz. This 10 kHz upper frequency limit was chosen for a number of reasons. Firstly, in-vivo studies performed on sounds recorded from human subjects with implanted valves have only ever shown sound components below 10 kHz. This inconsistency between in-vivo and in-vitro bandwidths can be accredited to the low-pass frequency characteristics of the chest thorax system [80]. At high frequencies (>10 kHz), sound attenuation due to the chest thorax structure is expected to render any valve signal undetectable. Another reason for setting this 10 kHz upper limit to the recorded metallic heart sounds is that in past studies that have considered frequencies above this point [73][81] no evidence has emerged to suggest these very high frequencies offer any diagnostic potential.

Figure 3.4 and 3.5 show the amplitude and phase response of channel 1 and channel 2 of the pre-processing hardware respectively.

---

**Figure 3.4:** Measured amplitude and phase response of pre-processing hardware channel 1.
Figure 3.5: Measured amplitude and phase response of pre-processing hardware channel 2.

It can be seen in Figure 3.4 that channel 1 exhibits a flat amplitude response and linear phase response over the 50 - 2000 Hz bandwidth of interest. In Figure 3.5 Channel 2 shows a similar desirable response over the 50 - 10000 Hz bandwidth of interest.
3.2 Phonocardiographic Transducer

Probably the most important stage of the data acquisition system is the initial conversion of the signal from vibration energy to an electrical signal. The integrity of this stage of data acquisition is essential as no amount of future processing can reinstate signal components rejected at this point. In the previous section there was a discussion of the desired frequency response of the recording system and Figures 3.4 and 3.5 illustrate the bandwidth over which the pre-processing hardware maintains signal integrity. It can be seen that the ideal microphone for use in the recording of both biological and metallic valves would have a flat frequency response from 50 - 10000 Hz.

The transducers available for the recording of PCG signals are almost all either air-coupled or direct-coupled devices. Of these two microphone types, the direct-coupled is the most desirable because of its low sensitivity to external noise and its elimination of a possible impedance mismatch between the sound source and the transducer. However, these direct coupled devices require a static application force to achieve good coupling which limits their useful bandwidth to 2-3 kHz.

There is no ideal solution to the phonocardiographic transducer requirements, hence a compromise must be found. A compromise is achieved by employing two alternative microphones for the two recording channels. Due to the limited bandwidth requirements of channel 1, a contact microphone with the desirable noise rejection and impedance characteristics was used. The Hewlett-Packard (21050A) contact microphone (mic 1 in Figure 3.3) was used together with the channel 1 hardware setup for the recording of PCG signals from subjects with native and bioprosthetic heart valves. The manufacturers quote a flat frequency response for this microphone between 0.02 Hz and 2 kHz, which is supported by the experimental results presented in [82][83]. Recording sounds from subjects with implanted mechanical prosthetic valves necessitates the wide bandwidth of channel 2 and the use of an air-coupled microphone. The Knowles BL 1994 air-coupled microphone (mic 2 in Figure 3.3) was used, together with the channel 2 hardware setup, for the recording of PCG signals from subjects with mechanical prosthetic valves. The response of this microphone is quoted as being flat from 20 Hz up to 10 kHz. To facilitate recording, the microphone was incorporated into a purpose-built lightweight plastic housing. The microphone housing provides a means of consistent application and reduces the influence of undesirable external noise. Figure 3.6 shows a schematic of the physical appearance of the combined microphone and housing unit.
Figure 3.6: Construction of the air-coupled microphone and purpose-built lightweight plastic housing.

From Figure 3.6, it can be seen that the microphone is held away from the chest surface by a cylindrical air cavity. In this setup, the cavity dimensions have an effect upon the frequency response of the microphone [84]. Dimensions of the plastic housing govern the dimensions of the contained cavity which in turn govern the resonance modes of the microphone unit. It can be shown that for a cavity span of 3 cm and a depth of 1 cm, resonance nodes occur above 10 kHz and hence, the integrity of the transducer frequency response is maintained.

3.3 Recording Protocol

Recordings were made with the subjects lying face up in a comfortable relaxed position. Both the contact and air-coupled microphones were held against the patient's chest with a rubber belt that passed around the patient's girth. Sounds due to the operation of the heart appear over almost all the surface of the chest but there are particular regions at which the operation of the aortic and mitral heart valves can be best detected. The "aortic area", at which, historically, the second heart sound has been best detected, can be defined as the area around the second right intercostal space, approximately 3 cm to the right of the upper sternal border. Similarly, the first heart sound has, in the past, been reported as having maximum intensity at the apex (behind the fifth left intercostal space, 8 - 9 cm from the midsternal line). Due to the variation in the size of heart and chest thorax structure from patient to patient, the position of these first and second sound areas is not constant. For this reason, the exact positioning of the transducer was determined by inspection of the trace, followed by
transducer repositioning until a point of maximum sound intensity was found. Figure 3.7 shows the general position of these two recording sites.

![Diagram of heart with microphone positions]

**Figure 3.7:** Microphone positioning for recording of aortic and mitral valve sounds.

For completeness, each subject's heart sounds were recorded four times. Each of these four recordings were performed with a different microphone/position setup. The four recording combinations were:

1. Channel 1, microphone 1, aortic area
2. Channel 1, microphone 1, mitral area
3. Channel 2, microphone 2, aortic area
4. Channel 2, microphone 2, mitral area

It should be noted that for each subject, only one of these recordings would probably be used for analysis, but at this early stage of the work it was felt necessary to make a complete record of each subject. This was particularly relevant when recording dysfunctioning subjects that would quite probably undergo corrective surgery, hence changing their normal/abnormal status.
Alongside the PCG record, a synchronous ECG recording was made to provide timing reference information. The ECG recorder was connected to the patient via self-adhesive electrode pads on each limb on an area of soft tissue approximately 10-15 cm from the hands and feet.

### 3.4 Patient Population

The population of subjects used in this work exhibit a range of valve types and conditions. The full set of recordings were collected over a two year period with an average of two subjects being interviewed per week. In total over 150 subjects were interviewed and over 1280 individual signals recorded. The subjects were recorded as they became available and a single visit to make recordings took on average three hours. Figure 3.8 shows the distribution amongst various valve types of the 100 subjects used in this study. It can be seen that in the case of native valve subjects there are, in total, 47 recordings. This native valve group is composed of two sub-groups: 21 normal valve subjects and 26 subjects with diagnosed abnormal valves. The abnormal subjects are further divided into 14 subjects with abnormal aortic valves and 12 subjects with abnormal mitral valves. Figure 3.8 shows a similar subdivision for bioprosthetic, metallic and before / after populations where appropriate. It should be noted that although the before / after population has only 16 subjects, each subject was recorded twice, once before surgery and once after surgery. Hence, the before / after population actually contains 32 sets of recordings.

![Diagram showing population distributions for recorded subject valve type and condition.](image-url)

**Figure 3.8:** Population distributions for recorded subject valve type and condition.
When attempting to characterise any property of a population by investigating a sample, a fundamental question arises: "How large should the sample be to produce a result representative of the whole population?". In an ideal situation, the whole population should be analysed. In accordance with this ideal, it may be stated that for the investigation to be successful the sample size should be as large as possible. The sample sizes shown in Figure 3.8, as in any survey, are practically limited by the availability of resources. It can be seen that significant populations were recorded for both native and C-E valves in normal and abnormal cases. Collection of significant samples for both valve types and in both states of function allows these populations to be used to investigate techniques for the classification of abnormal/normal valves. It was initially intended that a similar abnormal/normal study would be performed for metallic valves but it soon became clear that there were only a small number of dysfunctioning valves available for recording. The reasons for this unavailability of metallic valves was discussed in section 2.2.2.

In the case of bioprosthetic valves, there were a whole array of valves that had been implanted in patients in the Edinburgh area. To ensure consistency of results it was decided at an early stage that the investigation of bioprosthetic valves should concentrate on subjects with implanted C-E valves.

A significant population illustrated in Figure 3.8 is that of patients recorded before and after surgery. It was felt that by recording subjects before and after implantation of a mechanical prosthesis it would be possible to investigate the influence of two very different valve types on recorded sounds when under the influence of identical system conditions. It should be noted that a number of the "before surgery" subjects were also included in the population of subjects classed as having abnormal native valves.

3.5 Typical Results

Figure 3.9 shows typical PCG and ECG waveforms recorded from a range of the subject types illustrated in Figure 3.8.
Figure 3.9: Typical PCG and ECG traces recorded from subjects included in this study: (a) a full twelve second recording taken from a subject with normal native valves; (b) a 3 second segment of the trace shown in (a); (c) 3 seconds of a recording taken from a subject with a normal aortic C-E bioprosthetic valve; (d) 3 seconds of a recording taken from a subject with a normal aortic 23 mm Bjork-Shiley tilting disc valve; (e) 3 seconds of a recording taken from a subject diagnosed as suffering from severe aortic stenosis.
3.6 Summary

Described in this chapter were the hardware and associated procedures used to record heart sounds. Following a description of the basic structure of the system, a detailed illustration of pre-processing hardware was presented. Particular attention was paid to the justification of the choice of recording bandwidth to ensure high signal-to-noise ratios and hence, clear recordings. Together with the description of hardware, the recording procedures and patient populations were also illustrated. Finally, a number of typical ECG and PCG recordings have been given.

From the detailed description of equipment and methods, it is valid to conclude that the recorded sounds will be of a high quality and will provide no undesirable bias. This conclusion is supported by the integrity of the typical examples given in section 3.5. It can be seen from section 3.4 that over 100 subjects have been included in the survey. With such a large set of data it is envisaged that with further analysis it will be possible to answer the questions raised and objectives set in chapter 1.
Chapter 4

Time-Frequency Spectral Estimation

As already discussed, time-frequency signal analysis is a technique that offers a number of advantages over the usual time-domain and frequency-domain signal analysis methods. This chapter describes the theoretical and practical details of four alternative transforms that can be used to investigate the time-frequency nature of recorded heart sounds. The four transforms described are the short-time Fourier transform (STFT), the wavelet transform (WT), the Wigner distribution (WD) and the Choi-Williams distribution (CWD). Alternative descriptions of these four algorithms and other time-frequency transforms are presented in the tutorial paper by Hlawatsch and Boudreaux-Bartels [85]. After describing the theory behind each of these transforms a comparative analysis is performed of their application to test signals and a typical first heart sound. The chapter is concluded with the analysis of these test results to provide a decision as to the most suitable transform for the description of the heart sound data.

4.1 The Short-Time Fourier Transform (STFT)

In signal analysis few, if any, tools are as universal as the Fourier transform. It is used as the keystone of modern signal processing. The Fourier transform and its inverse are defined as follows.

\[ F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} \, dt \]  

(4.1)

\[ f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} \, d\omega \]  

(4.2)

Where \( F(\omega) \) is the Fourier transform of the signal \( f(t) \). Using the identity
the inverse transform can be described in terms of sine and cosine rather than complex exponentials:

\[ f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \left( \cos \omega t + j \sin \omega t \right) d\omega \]  

(4.4)

From this it can be seen that the Fourier transform \( F(\omega) \) of a signal is a function describing the contribution of sines and cosines to the construction of the original time-domain signal. The contribution of these so-called basis functions is limited in time only by the duration of the signal being analysed.

The time independence of the basis functions of the Fourier transform results in a signal description purely in the frequency-domain. The description of a signal as either a function of time or as a spectrum of frequency components contradicts our everyday experiences. The human auditory system relies upon both time and frequency parameters to identify and describe sounds. For example, to reproduce a piece of music, not only should all the correct notes be played but these notes should also be played in the correct order.

The Fourier transform is ideal for the analysis of stationary signals (signals whose statistical properties do not evolve with time). For the analysis of non-stationary signals a function is required that transforms a signal into a joint time-frequency domain. Such a description can be achieved using the well known STFT, which is an extension to the classical Fourier transform as defined by Gabor [3].

\[ STFT(\tau, \omega) = \int_{-\infty}^{\infty} s(t) \ g^*(t - \tau) \ e^{-j\omega t} \ dt \]  

(4.5)

This function can be described as the Fourier transform of the signal \( s(t) \), previously windowed by the function \( g(t) \) around time \( \tau \). As the window function is shifted in time over the whole signal and consecutive overlapped transforms are performed, a description of the evolution of signal spectrum with time is achieved. This method assumes signal stationarity over the limited window \( g(t) \). If the window is relatively short this assumption of local stationarity is often valid. A diagrammatical description of (4.5) is shown in Figure 4.1. As the window function \( g(t) \) is shifted in time, repeated Fourier transforms are performed on overlapped data sets. Arranging these transforms chronologically and on a common frequency axis provides a time-frequency description of the signal which is commonly called the signal spectrogram.
Figure 4.1: STFT sliding window analysis.

Figure 4.2 shows the result of applying the STFT, as described above, to a linear frequency modulated (LFM) test signal, Figure 4.2a. Figure 4.2b shows the Fourier transform of the LFM signal. Inspecting a signal as either a function of time or as its Fourier spectrum is inappropriate in many cases. Figure 4.2 illustrates that for a complete description of the non-stationary LFM signal, its time-frequency description (Figure 4.2c) is far more appropriate.

The introduction of a window into the Fourier transform, as in the STFT, allows the generation of a time-frequency description of a signal. The introduction of such a window, however, has a detrimental effect upon frequency resolution [86][87]. From Figure 4.1 it can be seen that choosing a short window will result in a transform exhibiting good resolution in time. A short window also results in a reduced number of samples used in the Fourier transform calculation which yields a reduced number of discrete frequencies that can be represented in the frequency-domain. With a reduced number of discrete frequency intervals, the ability of the transform to discriminate between sinusoids of different frequencies is significantly reduced.
The relationship between resolution in time and in frequency is a concept shared by many areas of science and is commonly referred to as the uncertainty principle [88]. If $\Delta t$ is the transform resolution in the time-domain and $\Delta f$ is the transform resolution in the frequency-domain, then two sinusoids will be discriminated only if they are more than $\Delta f$ apart in frequency or $\Delta t$ apart in time. The uncertainty principle [88] can be written as:
\[ \Delta t \times \Delta f \geq \frac{1}{4\pi} \]  

(4.6)

This equation implies that both time and frequency resolution cannot be made arbitrarily small, one must be traded for the other. The time and frequency resolution of the STFT is dependent upon the shape and length of the window function. Both these factors remain constant throughout an analysis and hence, the time-frequency resolution also remains constant throughout an analysis. This constant joint time-frequency resolution results in the STFT covering the time-frequency plane with a uniform array of resolution squares, as shown in Figure 4.3a. This trade off between time and frequency resolution can be illustrated with a few simple examples.

![Figure 4.3: Coverage of the time-frequency plane. The dimensions of the resolution squares represent minimum time and frequency intervals over which separate signals can be differentiated: (a) STFT; (b) WT.](image)

Figure 4.4 shows two simple test signals both 256 ms long and sampled at a rate of 2000 samples per second. The test signal in Figure 4.4a is comprised of two sinusoids of equal amplitude, one at 64 Hz and another at 192 Hz. The second test signal in Figure 4.4b contains a single sinusoid with a 64-sample gap, during which the signal is "switched off". Figure 4.5a shows the STFT of the test signal in Figure 4.4a using a relatively long window length of 128 samples and Figure 4.5b shows the STFT of the same signal with a shorter 32-sample window. These results clearly demonstrates how transform window length affects
resolution in frequency. The longer window length results in better frequency resolution and, as can be seen in Figure 4.5a, the two sinusoids are clearly resolved. The shorter window length results in reduced frequency resolution and this analysis, shown in Figure 4.5b, fails to resolve the two sinusoidal components.

Figure 4.6 shows the results of transforming the test signal in Figure 4.4b, again using a long 128-sample window and a short 32-sample window. As would be expected, applying the longer 128-sample analysis (Figure 4.6a) fails to resolve the signal gap but the short 32-sample analysis (Figure 4.6b) quite clearly resolves the position and length of the gap. Figure 4.5 and Figure 4.6 together illustrate the joint time and frequency resolution limitations of the STFT implied by the uncertainty principle in (4.6).

Although limited, the STFT is applicable to many problems where high resolution characteristics are not required. When signal composition warrants high resolution or signal composition is unknown, alternative techniques are required. The next section describes the WT as an alternative to the STFT for non-stationary signal analysis.

![Figure 4.4: Test signals: (a) 64 Hz and 192 Hz; (b) 128 Hz with a 64-sample gap.](image)
Chapter 4: Time-Frequency Spectral Estimation

Figure 4.5: STFT of the test signal in Figure 4.4a: (a) using a 128-sample window; (b) using a 32-sample window.

Figure 4.6: STFT of the test signal in Figure 4.4b: (a) using a 128-sample window; (b) using a 32-sample window.

4.2 The Wavelet Transform (WT)

Wavelet theory was initially proposed by the geophysicist J. Morlet and the theoretical physicist A. Grossmann [89]. Together with their fellow Frenchman, Y. Meyer, this "French school" developed the mathematical foundations of wavelets (ondelettes). At this stage wavelets were still very much in the realms of pure mathematics and, as such, concentrated
more on the theory than the application. The two American based researchers Daubechies and Mallat changed this by defining the connection between wavelets and digital signal processing [90].

As a time-frequency transform (the WT is more correctly described as a time-scale transform), the WT is a direct alternative to the STFT. Hence, the WT can be described as a modification and alternative to the STFT. Previously, the STFT (4.5) has been analysed as the Fourier transform of the windowed signal \( s(t)g(t - \tau) \), but it is just as valid to describe this function as the decomposition of the signal \( s(t) \) into the windowed basis functions \( g(t - \tau)e^{-j\omega t} \). The term "basis functions" refers to a complete set of functions that can, when combined as a weighted sum, be used to construct a given signal. In the case of the STFT these basis functions are complex sinusoids, \( e^{-j\omega t} \), windowed by the function \( g(t) \) centred around \( \tau \). Using this description it is possible to write a general equation for the STFT in terms of basis functions and signal as an inner product.

\[
STFT(\tau, \omega) = \int_{-\infty}^{\infty} s(t) g^*(t - \tau) dt
\]

For the STFT, the basis functions in (4.7) can be represented by \( g_{\tau,\omega}(t) = g(t)e^{-j\omega t} \). Figure 4.7a shows the real part of three such functions demonstrating the shape of typical STFT basis functions. These windowed basis functions are distinguished by their position in time \( \tau \) and their frequency \( \omega \). By mapping the signal onto these basis functions, a time-frequency description of the signal is generated.

The WT can also be described in terms of its basis functions, known as wavelets, using (4.7). In the case of the WT the frequency variable \( \omega \) is replaced by the scale variable \( a \) and generally the time-shift variable \( \tau \) is represented by \( b \). The wavelets are represented by \( h_{b,a}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t - b}{a}\right) \) where \( h^*(t) \) refers to the complex conjugate of \( h(t) \). Substituting this description into (4.7) gives the definition for the continuous wavelet transform (CWT).

\[
CWT(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h^*\left(\frac{t - b}{a}\right) s(t) dt
\]

From this equation it can be seen that the WT performs a decomposition of the signal \( s(t) \) into a weighted set of scaled wavelet functions \( h(t) \). In general the wavelet \( h(t) \) is a complex valued function, Figure 4.7b shows the real part of the Morlet wavelet [91] at three different levels of scale. Comparing the two sets of basis functions in Figure 4.7, it can be seen that the wavelets are all scaled versions of a common 'mother wavelet' while the basis functions of the STFT are windowed sinusoids.
Due to the scaling shown in Figure 4.7b, wavelets at high frequencies are of limited duration and wavelets at low frequencies are relatively longer in duration. This varying window structure is reflected in the coverage of the time-frequency plane by the WT, as shown in Figure 4.3b. These variable window length characteristics are obviously suited to the analysis of signals containing short high frequency components and extended low frequency components, which is often the case for signals encountered in practice. The WT employs a set of basis functions that are scaled versions of a single ‘mother function’. The ‘mother wavelet’ is constrained by a minimum of factors [92] and, as a result, the function can take many forms. Figure 4.8 shows a number of commonly used wavelets as cited in [90][91][93].

Probably the most frequently used wavelet is the Morlet wavelet shown in Figure 4.7b and defined as:

\[ h(t) = e^{j\omega_0 t} e^{-t^2/2} \]  

(4.9)

As can be seen the Morlet wavelet is a monotonic sinusoid windowed by a Gaussian function. The monotonic nature of this wavelet function results in a direct relationship between scale and frequency. This direct scale/frequency property of the Morlet wavelet makes it particularly attractive when applying the WT to the problem of time-frequency spectral estimation. Figure 4.9 shows both the magnitude and phase plots generated by applying the WT, using the monotonic Morlet wavelet, to the LFM test signal in Figure 4.2. Before describing
Chapter 4: Time-Frequency Spectral Estimation

the interpretation of Figure 4.9, it is necessary to first describe the conventions used to graphically display the results of the WT.

![Wavelet Examples]

**Figure 4.8:** Common wavelets used by previous researchers.

The CWT as a time-scale transform has three dimensions. In this thesis the three dimensions are represented on a log(a),b half plane. The log(a) axis (scale) faces downwards and the b axis (time-shift) faces to the right. The respective intensity of the transform at points in the log(a),b half plane is represented by grey level intensity. These conventions allow a standard plot of time evolution from left to right and decreasing frequency from top to bottom, as shown in Figure 4.10. Using a log scale axis allows a large range of scales to be simultaneously displayed. Analysis of a large range of scale parameters is often essential for example, audible sounds have a ten octave scale range.
Figure 4.9: The WT of the standard LFM test signal: (a) LFM test signal; (b) WT magnitude plot; (c) WT phase plot.

Figure 4.10: Time-scale half plane used to display both magnitude and phase of the WT result.
The phase plot of a transformed signal does not show the same localisation as its equivalent modulus plot. Signal phase is a relative measure of complex and real components and, unlike modulus, it is not governed by localisation of signal intensity. For this reason the phase plot generated by the transformation of a real signal invariably covers the whole \( \log(a),b \) plane. Such a plot is often difficult to interpret and, for this reason, the phase is selectively displayed. The phase is not displayed if the transform modulus drops below a predetermined cut-off value. Selective display of the phase plot is an aid to interpretation and for maximum clarity individual plots require individual thresholds.

The transformation of the LFM test signal in Figure 4.9 clearly demonstrates these graphical conventions. By comparing the modulus plots of the LFM test signal for both STFT and WT, Figure 4.2c and Figure 4.9b respectively, it can be seen that both transforms produce similar results. The major difference between these two plots is that the WT results in a time-scale interpretation rather than the time-frequency description generated by the STFT. Inspecting the phase plot in Figure 4.9c it can be seen that clear lines of constant phase are present. The separation of these constant phase lines is directly dependent upon signal frequency and independent of wavelet choice [94]. By measuring the separation of these phase lines signal frequency can be estimated and the relationship between the STFT time-frequency and the WT time-scale representation becomes apparent.

The STFT was originally introduced as a transform that performs signal analysis using a ‘sliding’ time window, the WT can also be described in this context. The distinct difference between each transform is that the STFT employs time windows of constant length over all frequency values whereas the CWT, due to the inclusion of the scale term \( a \) into the window function, employs time windows that decrease in length as frequency increases.

By employing scaled window functions the WT does not overcome the uncertainty principle, but by employing variable window lengths and hence, variable resolution, an increase in performance may be achieved. Figure 4.5 and Figure 4.6 demonstrated the resolution limitations imposed upon the STFT by the uncertainty principle. Figure 4.11 shows the magnitude and phase plots generated by applying the WT to the two test signals of Figure 4.4. It can be seen from Figure 4.11 that in both test cases the WT has clearly separated the signal components.
Figure 4.11: Wavelet transform of the test signals in Figure 4.4: (a) the magnitude and phase plot of the WT of the two-sinusoid test signal (Figure 4.4a); (b) the magnitude and phase plot of the WT of a sinusoid-with-gap test signal (Figure 4.4b).

An important feature of Figure 4.11b is the characteristic upturned V's that are present at the discontinuities in the interrupted sinusoid. It can be seen that the positions of the signal discontinuities are clearly visible due to localisation in time at high frequency. This localisation about signal discontinuities is an important characteristic of the WT, that can be used in signal interpretation [95].

4.2.1 Implementation of the WT

The areas to which the WT is applied are varied and, as such, the WT has previously been implemented in a number of ways. This section describes the two most common manifestations of the WT: the true discrete wavelet transform (DWT) and the discrete version of the CWT.
The discrete WT

To generate a scalogram plot comparable to the STFT spectrogram (Figure 4.2), a discrete version of the original CWT equation must be considered:

\[ CWT[iT_s, a] = T_s \frac{1}{\sqrt{a}} \sum_{n=0}^{N-1} h\left[\frac{n-i}{a}\right] s[nT_s] \]  

(4.10)

Where \( N \) is the number of samples in the signal and \( T_s \) is the sampling interval. From this equation, it can be seen that at all values of the scale parameter \( a \), a full set of \( n \) samples are generated, i.e., no time-domain subsampling takes place between octaves. It can also be seen that \( a \) is unconstrained; that is, it can take any discrete value allowing the octave analysis to be subdivided. These octave subdivisions are often called voices. The values taken by the scale parameter \( a \) at discrete octave and voice levels of scale can be expressed as the series shown below [96].

\[ a = 2^{j+m/M} \]  

(4.11)

where \( j \) is the octave number, \( m \) is the voice number and \( M \) is the number of voices per octave. Figure 4.12a shows the time-scale grid generated using a number of voices per octave. The discrete version of the CWT results in a magnitude and phase plot with \( N \) samples along the time-shift axis (where \( N \) is the number of signal points) and \( J \times M \) samples along the scale axis (where \( J \) is the number of octaves). Using a suitably dense grid allows a scalogram with similar appearance to the typical STFT spectrogram to be produced.

\[ \text{time-shift, } b \]

(a)

log (scale, a)

(b)

log (scale, a)

\[ \text{time-shift, } b \]

Figure 4.12: Sampling of the time-scale plane: (a) the discrete CWT; (b) the DWT.
The true discrete wavelet transform (DWT)

The discrete version of the CWT (4.11), implemented with a number of voices per octave exhibits a large degree of redundancy. The redundancy in scale can be removed if a single analysis per octave is performed and redundancy in the time-shift variable can be reduced to a minimum by subsampling the signal after each octave analysis. This time and scale subsampling results in the time-scale grid shown in Figure 4.12b. This redundancy-free DWT is implemented as the tree structure shown in Figure 4.13.

![Diagram of the DWT implemented as subband decomposition using a filter bank tree.](image)

**Figure 4.13:** Block diagram of the DWT implemented as subband decomposition using a filter bank tree.

Implementing this filter tree structure results in a bank of contiguous bandpass filters spread logarithmically over frequency as shown in Figure 4.14.
The frequency-domain coverage of the DWT shown in Figure 4.14 is identical to that exhibited by subband octave signal decomposition techniques. Also, the tree structure of Figure 4.13 is identical to that for signal subband decomposition. When implementing the filters in Figure 4.13 it is desirable that together they achieve complete spectral coverage. It is clear that it is impossible to use brickwall filters but in subband coding techniques an approximation to complete coverage is achieved using quadrature mirror filters (QMF) [97]. A similar solution is achieved in the DWT by placing a number of constraints upon the wavelet functions for it to exhibit QMF characteristics at each level of frequency octave.

The wavelet coefficients created by the above band pass filter bank structure are a true subsampled set of the continuous wavelet transform (4.8). Also, the wavelets used as the basis functions for this DWT form an orthonormal set. Representing the wavelet set as $h_{b,a}(t)$, orthonormality is defined by

$$\int h_{b,a}(t) h_{b',a'}^*(t) \, dt = \begin{cases} 1 & \text{if } b = b' \text{ and } a = a' \\ 0 & \text{for all other cases} \end{cases} \quad (4.12)$$

Orthonormality implies that no information redundancy is present. Orthonormality also implies that any finite signal can be represented as a weighted sum of the basis functions and, conversely, that the signal can be perfectly reconstructed from a full set of weighted functions. Figure 4.15 shows two alternative representations of the DWT of the LFM test signal shown in Figure 4.9a.
Chapter 4: Time-Frequency Spectral Estimation

Figure 4.15: DWT of the signal in Figure 4.9a: (a) DWT coefficients displayed sequentially; (b) DWT coefficients displayed in the time-scale plane.

Figure 4.15a shows the DWT coefficients sequentially arranged and Figure 4.15b shows the two-dimensional plot of the magnitude of the DWT coefficients in the time-scale plane. It can be seen that Figure 4.15b bears a resemblance to the continuous result (Figure 4.9b) but does not have the same clear continuous visual appearance. In spite of this lack of visual clarity the DWT, when used with an orthonormal wavelet set, has distinct advantages. These advantages include efficient signal decomposition and perfect signal reconstruction via an inverse transform.

Computation load

Although the direct implementation of the algorithm, as above, allows a clearer signal representation than the octave band DWT, this results in a higher computational load. The DWT implemented as the tree structure in Figure 4.12 is efficient. It utilises common filter operations and performs dilation via simple signal decimation. In practice $L$ sample points analysed over $J$ octaves require $2L(1 - 2^{-J})$ multiplications/point and $2(L - 1)(1 - 2^{-J})$ additions/point [96]. In contrast, using the discrete CWT (4.10) to produce a grid with $M$ voices per octave, as in Figure 4.14b, requires considerably more operations, $2LM$ multiplications/point and $2MJ(L - 1)$ additions/point [96]. If a speech signal lasting a few seconds
were to be transformed, a range of typically ten octaves along the scale axis and 8,000 time-domain samples per second would have to be analysed. To generate a clear plot as many as eight voices per octave might be used. Considering these typical values it can be seen that to generate transforms with a high degree of detail over extended periods would require an extensive amount of computation. This fact has been noted by a number of researchers and has lead to the development of a number of reduced computation algorithms for calculating WTs [96][98][99].

One such algorithm is based upon the FFT. By analysing the CWT in the Fourier domain the basic convolution operation of the CWT can be achieved via simple multiplication operations. Writing the CWT in the Fourier domain gives.

\[ F\{CWT(b, a)\} = \frac{1}{\sqrt{a}} H^*(\frac{\omega}{a}) S(\omega). \]  

(4.13)

Where \( F\{CWT(b, a)\} \) is the Fourier transform of the continuous wavelet transform. \( H^*(\omega) \) is the Fourier transform of the wavelet and \( S(\omega) \) is the Fourier transform of the input signal. This equation can be used as the basis of an FFT-based fast WT. Figure 4.16 shows a graphical representation of (4.13).

Figure 4.16: Block diagram of the FFT-based fast WT.

Figure 4.16 shows that the WT, after pre-calculation of the FFT of the wavelet and signal, is implemented via repeated scale, multiply and inverse FFT (IFFT) operations. This simple structure results in a complexity of \([6 + (n - 3) + 2/L]MJ\) multiplications/point and \([6 + (3n - 5) + 4/L]MJ\) additions/point. Where \( n = \log_2(2L) \) and split-radix [100] implementation of the IFFT has been assumed.
Practically, the analysis of $2^{14}$ samples of a signal (approximately 2 seconds at 8 kHz sampling rate) is over 700 times less complex when calculated using the fast WT algorithm rather than the direct implementation of the WT. Clearly the FFT-based fast WT and other equivalent fast algorithms provide significant computational savings. This not only allows the analysis of significant sample lengths but also real time analysis becomes more realistic.

### 4.3 The Wigner Distribution (WD)

The subject of time-frequency signal description is motivated by a need for a function in two variables that will give the fraction of total energy of a signal at a time $t$ and frequency $f$. In light of this aim the WD (4.14) was first proposed by Ville [101] and hence is commonly referred to as the Wigner-Ville distribution.

$$W(t, f) = \int_{-\infty}^{\infty} s^\ast(t - 1/2\tau) s(t + 1/2\tau) e^{-j2\pi ft} d\tau \quad (4.14)$$

Where $s^\ast(t)$ is the complex conjugate of the signal of interest $s(t)$. The WD is particularly useful because it satisfies a great number of mathematical properties and indeed a great number of desirable time-frequency properties. As described by Jeong and Williams [102] the WD satisfies eight of the eleven desirable time-frequency distribution properties while maintaining a very high level of joint time-frequency resolution. These exceptional qualities of the WD can be seen when analysing a chirp signal. Analytically we have for the signal $s(t)$:

$$s(t) = \left(\frac{\alpha}{\pi}\right)^{\frac{1}{4}} e^{-\frac{1}{2}a^2 + \frac{1}{2}b^2 + j2\pi f_0 t} \quad (4.15)$$

Where $\alpha$ and $\beta$ define the rate of change of amplitude and phase respectively in the chirp signal and $f_0$ is the chirp's initial frequency. Substituting this chirp signal into (4.14) gives:

$$W(t, f) = 2e^{-a^2} e^{-A \pi^2 (f^2 + \beta t^2 + 2\pi f_0 t)} \quad (4.16)$$

The significance of this result is that it is positive at all times and lies along the line $f = f_0 + \beta t/2\pi$, which is the derivative of the phase. Considering this example the performance of the WD appears almost ideal but, as stated earlier, in the general case the WD fails to meet two of Jeong's desirable time-frequency characteristics [102]. These two characteristics are non-negativity and reduced interference, the WD satisfies the first property in only one signal case, that given in (4.15). The second property is also not violated in the case of (4.15) if the analytic form of the chirp signal is evaluated. The analytic form of a signal is generated by transforming the complex signal into the frequency-domain and discarding the negative frequency components. This results in a signal whose real part is the original signal.
and the imaginary part is the Hilbert transform of the real signal. There are two main reasons for using the analytical signal rather than the true complex signal. The first is that for a physical signal negative frequencies have no physical meaning and hence, need not be considered. The second reason is clearly demonstrated in Figure 4.17 where the WD of a LFM chirp signal, for both analytic and true complex signals is illustrated. Both signals are of length 512 points and sampled at 512 Hz. The signal rises linearly in frequency from 10 Hz to 100 Hz between 64 ms and 960 ms respectively and is zero at all other times.

In Figure 4.17 it can be seen that the WD of the true complex signal has a significant component at d.c. that follows the time evolution of the time-domain signal. This d.c. term is a result of interference between positive and negative frequency components, hence, when transforming the analytic signal this interference does not occur (Figure 4.17b).

![Figure 4.17: WD of a LFM chirp: (a) true complex signal; (b) analytic signal.](image)

The WD of the analytic LFM chirp signal shows an almost ideal result, very good localisation in time and frequency with an accurate representation of instantaneous signal frequency. As already mentioned, the WD of a single analytic chirp signal is a special case and is, in fact, the only situation in which all eleven of Jeong’s desirable time-frequency properties are demonstrated. In the general case both undesirable properties of negative values and significant interference terms are grossly violated. The occurrence of these two undesirable
Chapter 4: Time-Frequency Spectral Estimation

properties can be illustrated by considering a signal that is a sum of two parts:

\[ s(t) = s_1(t) + s_2(t) \]  \hspace{1cm} (4.17)

The WD of the composite signal is given by:

\[ W(t, f) = W_{11}(t, f) + W_{22}(t, f) + W_{12}(t, f) + W_{21}(t, f) \]  \hspace{1cm} (4.18)

where

\[ W_{nm}(t, f) = \int_{-\infty}^{\infty} e^{-j2\pi f' \tau} s_n^*(t - 1/2\tau) s_m(t + 1/2\tau) \, d\tau \]  \hspace{1cm} (4.19)

It can be seen that the WD of a multicomponent signal is not just the sum of the distributions of the individual signals but it also contains joint terms. Considering a two sinusoid component signal \( s(t) \) we have:

\[ s(t) = A_1 e^{j2\pi f_1 t} + A_2 e^{j2\pi f_2 t} \]  \hspace{1cm} (4.20)

Substituting this signal into the WD (4.14) we have:

\[ W(t, f) = A_1^2 \delta(f - f_1) + A_2^2 \delta(f - f_2) \]

\[ + 2A_1A_2 \delta(f - \frac{1}{2}(f_1 + f_2)) \cos[2\pi(f_2 - f_1)t] \]  \hspace{1cm} (4.21)

The first two terms in (4.21) represent a concentration in the time-frequency domain for all time at frequencies \( f_1 \) and \( f_2 \), as is desirable for the dual frequency signal. There are also additional terms at frequency \( \frac{1}{2}(f_1 + f_2) \) modulated in time by \( \cos[2\pi(f_2 - f_1)t] \). Obviously this interference component has both positive and negative values. Hence, the WD of this multicomponent signal has interference terms and negative values. Figure 4.18 shows the magnitude of the WD generated for one second of a signal comprising of two frequency components at 8 Hz and 24 Hz. It can be seen that together with the expected frequency components there is an additional term at \( \frac{1}{2}(f_1 + f_2) \) (16 Hz) that clearly oscillates as a function of time. A more detailed description and a number of practical examples of interference terms generated by the WD of multicomponent signals are presented in the tutorial paper by Hlawatsch and Boudreaux-Bartels [85]. The above analytic analysis of a two component signal indicates that the interference term will take negative values, but as Figure 4.18 is the magnitude of the result the negative terms are reflected into positive results and the interference oscillates at twice the expected frequency.
Chapter 4: Time-Frequency Spectral Estimation

Figure 4.18: WD of a dual sinusoid signal.

The appearance of interference terms can be explained if we consider the structure of the WD equation and its behaviour when used to analyse a simple illustrative signal [103]. It is possible to describe (4.14) as a function that at time $t$ produces the summation of the result of multiplying a delayed version of the signal with a future version of the signal. The positive and negative shifts in the signal versions are equal. This operation on the signal can be visualised as a folding of the signal at the time of interest. If the signal is folded at time $t$ and any of the signal to the left of the fold overlaps the signal to the right of the fold then the WD result is non zero. To illustrate this explanation consider the finite signal with non zero components between $t_1$ and $t_2$:

Folding the signal about a point to the left of $t_1$ will result in no overlap which means the WD will be zero at this point in time. Folding the signal at any point between $t_1$ and $t_2$ will result in an overlap and hence the WD at this time will take a non zero value. Considering this result it can be seen that for this simple finite signal the WD will be zero at all points up to $t_1$ and past $t_2$ and will have non zero values when the signal is non zero. From this example the WD seems to behave very well, but as illustrated in (4.15) the WD looses its attractive properties when more complex examples are considered. Consider a signal with two
Chapter 4: Time-Frequency Spectral Estimation  

It can be seen that folding the signal between $t_1$ and $t_2$ or between $t_3$ and $t_4$ results in the WD taking a desirable non-zero value. In comparison, using $t_f$ as the folding point in the new signal, it can be seen that even though the signal is zero at this point the WD will take on a non-zero value. As has already been discussed, the appearance of these WD terms where true signal energy does not exist is highly undesirable.

The above argument has been based around the time-domain description of a signal but it can also be used to argue a completely analogue result in the frequency-domain. As was shown in Figure 4.18 the analysis of a signal with multiple frequency components produces non-zero WD components at points where signal energy does not truly exist. Expressing the WD in terms of the signal spectrum gives:

$$W(t, f) = \int_{-\infty}^{\infty} S^*(\omega + 1/2\theta) \ S(\omega - 1/2\theta) \ e^{-j2\pi f\theta} \ d\theta$$

(4.22)

It can be seen from (4.22) that the spectral description of the WD has the same delay summation form as the time-domain description (4.14). This duality of the time and spectral representations means that the illustration of interference terms described above, for the time-domain, is equally valid for the frequency content of a signal. As a result of this joint time and frequency interference behaviour, the analysis of signals with multiple time and frequency components results in combined time and frequency interference terms.

4.3.1 Strengths and weaknesses of the WD

The consideration of strengths and weaknesses is particularly apt for the WD due to its conflicting performance as a time-frequency descriptor. On one hand its localisation in time and frequency is almost ideal, Figures 4.17 and 4.18 both showed clear components at the desirable frequency and time slots for the signals being analysed. This exemplary performance is contrasted by the appearance of undesirable cross-terms. In a time-frequency transform of a real signal the negativity of these cross-terms is unexplainable with no logical interpretation.

66
in relation to the analysed signal. The second disadvantage of these cross-terms is that they indicate signal activity and energy concentration where one would not expect it. The dual sinusoid signal in Figure 4.17 was generated as the sum of two distinct sinusoid signals and as such, the ideal time-frequency distribution would indicate two unique lines across the resultant plot. In contrast to this ideal, the time frequency distribution shows three distinct lines of activity. In this simple case it is possible to easily distinguish between the true distribution and the interference. If a more complex and, more importantly, an unknown signal were analysed it is not unreasonable to assume that interference signals will become indistinguishable from the true signal components.

The WD and its application to problems is very much limited by the cross-term problem. For this reason extensive work has been carried out to try and remove and reduce the problematic cross-terms [102][104]. The next section details these new ideas, together with a new time-frequency distribution that addresses the cross-term problem.

4.4 The Choi-Williams Distribution (CWD)

After the discovery of the WD a number of other distributions with various advantages were developed [105][106][107]. These distributions showed common desirable properties, such as satisfying the marginals and obviously all present a reasonable time-frequency description of a signal. The appearance of these seemingly valid transforms all with alternative derivations and valid performances created a confusing situation. This confusion was all overcome when Cohen [108] presented his unified theory for time-frequency distributions. Cohen proposed that there were, in fact, an infinite number of possible transforms that could be derived from the general formula:

\[
P(t, f) = \int \int \phi(\theta, \tau) s^*(u + 1/2\tau) s(u - 1/2\tau) e^{-j2\pi(\theta + f \tau - \theta u)} \, d\theta \, d\tau \tag{4.23}
\]

Where \(s^*(t)\) is the complex conjugate of the signal of interest \(s(t)\) and \(\phi(\theta, \tau)\) is referred to as the kernel function. The formula shows a resemblance to the WD with the same bilinear structure, the \(s^*(u + 1/2\tau)\) and \(s(u - 1/2\tau)\) terms. This would suggest that in general the same folding argument as used in section 4.3 can be used to show the appearance of cross-terms in this general case. The important feature of (4.23) is the kernel function \(\phi(\theta, \tau)\) [103]. By choosing different kernel functions the different distributions previously developed, Wigner [101], Page [106], Rihaczek [107] and many new ones can be constructed. The distinct advantage of the kernel notation is that properties of the kernel determine and, in fact, can be used to control the properties of the distribution. As mentioned above the general distribution
Chapter 4: Time-Frequency Spectral Estimation

suffers from cross-terms due to the common positive-negative delay structure of the formula, but it is possible to use kernel functions that reduce the significance of these interference terms.

The satisfaction of desirable distribution properties has driven the development of various distributions with one in particular having been developed to address the cross-term problem. The distribution in question was first studied by Choi and Williams [104]. They named the new distribution the exponential distribution after its kernel function but future work has referred to it as the CWD. Choi and Williams [104] argued that instead of removing or filtering the cross-terms from the WD it was more useful to look for distributions which produced reduced interference terms. The main argument behind this being that the methods devised for removing or reducing cross-terms resulted in a final distribution that did not conform to the distribution desirable properties [102]. The Choi-Williams kernel is:

$$\phi(\theta, \tau) = e^{-\phi^2 \tau^2 / \sigma}$$  \hspace{1cm} (4.24)

Where $\sigma$ is a constant. Substituting this into the generalised distribution (4.23) we obtain the CWD:

$$P(t, f) = \int \int \sqrt{\frac{\pi \sigma}{\tau^2}} \ s^*(u - 1/2\tau) \ s(u + 1/2\tau) \ e^{-\pi^2 \sigma(u-t)^2 / \tau^2} \ e^{-j2\pi tf} \ \text{dud} \tau$$  \hspace{1cm} (4.25)

The implementation of this and, in fact, any distribution based on Cohen's general class (4.23) has three basic components. Breaking this general equation into three steps we have:

$$P(t, f) = \int \int r(u - t, \tau) \ e^{-j2\pi ft} \ (sac) \ \text{dud} \tau$$  \hspace{1cm} (4.26)

Where

$$r(u, \tau) = \int \phi(\theta, \tau) \ e^{j2\pi u \theta} \ \text{d}v$$  \hspace{1cm} (4.27)

$$(sac) = s^*(u - 1/2\tau) \ s(u + 1/2\tau)$$  \hspace{1cm} (4.28)

Where the term $(sac)$ is know as the symmetric time-varying autocorrelation of the signal $s(t)$. Inspecting this breakdown of the Cohen general time-frequency representation we have a simple three step implementation of any distribution.

1/ Calculate $r(u, \tau)$ and $(sac)$.

2/ Convolve with respect to time $r(u, \tau)$ and $(sac)$.

3/ Column FFT the result of 2.
This simple three step implementation can be used with various kernels. In the previous section we described the WD. As the WD is a member of Cohen's general class it has a kernel function that can be used in the above implementation. For the WD we have:

\[
\phi(\theta, \tau) = 1
\]  

(4.29)

Comparing this to the Choi-Williams kernel (4.24) we can see that the Wigner kernel is a special case of the Choi-Williams kernel. If the constant \( \sigma \) in (4.24) (referred to from this point as the Choi-Williams kernel spread) is large then the kernel tends to 1 for all values of \( \theta \) and \( \tau \). Figure 4.19 shows a three-dimensional plot of the Choi-Williams kernel with various values for the spread factor.

![Figure 4.19: Choi-Williams kernel function at different spread factors: (a) spread factor = 1; (b) spread factor = 10; (c) spread factor = 100; (d) spread factor = 1000.](image)

It can be seen that at high values of spread factor (Figure 4.19d) the CWD tends towards the WD and the distribution generated for a dual sinusoid signal will appear as for the WD.
shown in Figure 4.18. In accordance with the initial objective of Choi and Williams' work [104], as the spread factor is reduced the interference terms become less significant. Figures 4.20a and 4.20c show the CWD of a dual impulse signal at a high and low value of spread factor respectively. The dual impulse signal contains 64 points all of which are zero apart from two unit impulses at point 24 and 40. Figures 4.20b and 4.20d show a frequency cross section of the distributions at a normalised frequency of 0.5.

Figure 4.20: CWD of a dual impulse signal: (a) full distribution result at a high value for spread factor; (b) cross section of the distribution in (a) at a normalised frequency of 0.5; (c) distribution at a low spread factor; (d) cross section of the distribution in (c).

The results in Figure 4.20 show that as the spread factor is lowered, the cross-terms in the distribution result become reduced in magnitude and spread in time. Although the cross-terms are not strictly reduced in energy the effect of spreading the terms results in the cross-terms having a reduced effect upon the distribution at individual positions in the time
frequency plane. The behaviour of these cross-terms, for the CWD of chirp and sinusoid signals at various spread factors, is fully documented in Choi and Williams’ original paper [104].

Figure 4.20 shows how reducing the Choi-Williams kernel spreading factor results in a reduction in the significance of undesirable cross-terms. It can also be seen that reducing the spreading factor results in a spreading and loss of clarity of the desirable distribution terms. Hence, we have a trade-off for the CWD: Reduction of the influence of undesirable cross-terms using a reduced kernel spread factor results in a fall in the resolution of the distribution’s desirable self terms.

An important question now arises: what value do we give the Choi-Williams kernel spread factor to allow good resolution while significantly reducing the distribution cross-terms? An attempt to answer this question is presented by Jeong and Williams [102] by defining a modified class of distributions known as reduced interference distributions (RIDs). In their paper [102] Jeong and Williams presented eleven distribution properties that were deemed to be desirable for good time-frequency signal description. These are:

**P0.** Non-negativity: \( P(t, \omega; \phi) \geq 0 \ \forall \ t, \omega. \)

**P1.** Realness: \( P(t, \omega; \phi) \in R. \)

**P2.** Time shift: \( g(t) = f(t - t_0) \Rightarrow P_g(t, \omega; \phi) = P_f(t - t_0, \omega; \phi). \)

**P3.** Frequency shift: \( g(t) = f(t)e^{j\omega_0t} \Rightarrow P_g(t, \omega; \phi) = P_f(t, \omega - \omega_0; \phi). \)

**P4.** Time marginal: \( 1/2\pi \int P(t, \omega; \phi) d\omega = f(t) f^*(t). \)

**P5.** Frequency marginal: \( \int P(t, \omega; \phi) dt = F(\omega) F^*(\omega). \)

**P6.** Instantaneous frequency: \( \left[ \int \omega P(t, \omega; \phi) d\omega / \int P(t, \omega; \phi) d\omega \right] = \omega_i(t). \)

**P7.** Group delay: \( \left[ \int t P(t, \omega; \phi) dt / \int P(t, \omega; \phi) dt \right] = t_g(\omega). \)

**P8.** Time support: \( f(t) = 0 \ \text{for} \ |t| > t_c \Rightarrow P(t, \omega; \phi) = 0 \ \text{for} \ |t| > t_c. \)

**P9.** Frequency support: \( F(\omega) = 0 \ \text{for} \ |\omega| > \omega_c \Rightarrow P(t, \omega; \phi) = 0 \ \text{for} \ |\omega| > \omega_c. \)

**P10.** Reduced interference.
Where $P(t, \omega; \phi)$ is a general distribution in time and frequency implemented using the kernel function $\phi$. $F(\omega)$ represents the Fourier transform of the signal $f(t)$ and $F^*(\omega) / f^*(t)$ represent the complex conjugate of the frequency / time-domain signals. Properties P0 and P1 are self explanatory and are desirable to allow clear interpretation of the distribution result. Time and frequency shift properties (P2, P3) define a linear relationship between time / frequency shift in the signal and the resultant time / frequency shift in the distribution result. The marginal properties P4 and P5 indicate that summation of the distribution along the time axis or frequency axis should result in the signal spectrum or time-domain description respectively. As an extension to the marginal properties, P6 defines a relationship between the summed frequency distribution and signal instantaneous frequency. P7 defines a relationship between the summed time distribution and group delay. The support properties P8 and P9 indicate that it is desirable for the distribution to be concentrated about the points where signal components exist. These two properties can be loosely described as an expression of the need for high resolution. Properties P0-P9 are desirable properties that have been considered by previous researchers, but P10 was added by Jeong and Williams [102]. Jeong and Williams defined a new class of distributions, reduced interference distributions (RIDs), as time-frequency descriptors that satisfy properties P1-P10. Note, P0 is left out of this definition because satisfaction of all eleven properties P0-P10 is a problem yet to be solved. The distributions already considered (STFT, WD and CWD) all satisfy some of these properties. Table 4.1 shows the combination of properties satisfied for each of these transforms.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>P0</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>STFT</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Wigner</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Choi-Williams</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 4.1: Desirable properties satisfied by time-frequency distributions

It can be seen that the STFT satisfies the illusive P0 and the new property P10. The STFT fails to satisfy properties P4-P9 because of its lack of resolution, as already demonstrated in section 4.1. The WD performs very well by satisfying P1-P9 but as already demonstrated, P10 is violated to an extreme that the application of the WD to multicomponent signals is problematic. The CWD meets the new property P10 but in the general case fails to meet P8 and P9. In spite of this it is possible, by truncating the Choi-Williams kernel, to force the...
CWD to satisfy the support properties (P8, P9) and still maintain the satisfaction of all other properties apart from P0. The Choi-Williams kernel (4.24) can alternatively be expressed as a function of time:

\[
h(t) = \frac{1}{\sqrt{2\pi}\alpha} e^{-(t/2\alpha^2)}
\]  

(4.30)

Converting this in terms of \(\theta\) and \(\tau\), we Fourier transform this description and replace \(\theta\) by \(\theta\tau\). i.e.

\[
H(\theta) = \int h(t)e^{-j\theta t} dt
\]  

(4.31)

\[
\phi(\theta, \tau) = H(\theta\tau)
\]  

(4.32)

It can be seen that the spread factor \(\sigma\), shown in (4.24), is equivalent to \(\frac{1}{\alpha}\) in the time-domain description of (4.30). Three time-domain representations of the Choi-Williams kernel (4.30), for varying values of spread factor, are shown in Figure 4.21.

![Figure 4.21](image)

**Figure 4.21:** Time-domain representation of the Choi-Williams kernel at different time spread factors: (a) time spread factor 2; (b) time spread factor 4; (c) time spread factor 10.

The distribution is forced to comply to the support properties, P8 and P9, by truncating the time function of (4.30) using the window \(w(t)\). The new time description of the kernel becomes:
Chapter 4: Time-Frequency Spectral Estimation

\[ h(t) = \frac{1}{\sqrt{2\pi \alpha}} e^{-t^2/(2\alpha^2)} w(t) \]  

(4.33)

Where \( w(t) \) is a symmetrical window function with significant values limited to the range \(-0.5 \leq t \leq +0.5 \) [102]. From Figure 4.21 it can be seen that if the time spreading factor \( \alpha \) is large enough the above truncation is approximately satisfied, i.e. the time-domain Choi-Williams kernel approximates to zero when \(-0.5 \geq t \geq +0.5 \). As mentioned by Jeong and Williams [102], if spread factor for the CWD distribution is set low enough then properties P8 and P9 will be approximately satisfied. For this reason it was decided not to window the Choi-Williams kernel but to limit the time-domain kernel description such that:

\[ h(t) = 0 \quad \text{for} \quad |t| > 0.5 \]  

(4.34)

It was decided to use a value of spread factor such that:

\[ h(t) = \frac{h(0)}{100} \quad \text{for} \quad |t| = 0.5 \]  

(4.35)

The value of spread factor \( \alpha \), defined in (4.30), that satisfies (4.35) is 0.165. The kernel function generated using this spread factor is shown in Figure 4.22.

\[ h(t) \]

\[ t \]

Figure 4.22: Time-domain representation of the Choi-Williams kernel function with optimal spread factor \( (\alpha = 0.165) \).

This choice of spread factor allows the highest possible resolution whilst closely approximating the desirable properties P8 and P9. All future applications of the CWD shown in this chapter.
thesis are made using this optimised kernel function.

4.5 Data Display Tool

Before describing the analysis of results using the transforms described above, it is important that a description of the tool used to extract time-frequency information from the resultant plots is given. A number of tools are available for the display of two-dimensional images that provide a clear plot of results, but it was decided that for this specific application a tailor-made display and analysis tool was desirable. In particular, it was felt that by devising a tailor-made tool, it would be possible to control closely the display parameters and prevent any bias due to the display format used. Also, due to the extensive amount of data to be analysed it was felt productive to develop a tool with features biased towards the analysis of heart sounds in the hope of reducing the amount of time required for data analysis. The tool described here was developed in the C programming language using X-view macro functions to control window and graphical operations.

The display used is a colour display that represents linear changes in image intensity by a corresponding change in colour from red to green to blue. This tri-colour display provides the advantage that an initial overview of the signal intensity can be performed. In general, signal components that appear red are low intensity, green components are medium intensity and blue components are high intensity. A more detailed analysis can be achieved by inspecting the component colour closer and estimating where exactly it lies in the whole red-green-blue colour spectrum. Figure 4.23 shows a display of the actual screen appearance of the tool, including control panel and displayed image.

As shown in Figure 4.23 the tool’s control panel has a number of features. These features can be broken down into two categories, image display and image analysis. The image display features span from the plot button to the reverse colour button and are self explanatory operatives for plotting, loading, printing, colouring and resizing the data image. The features shown below the reverse colour button are the image analysis features that have been included in the tool to facilitate the analysis of the 2D time-frequency images of recorded heart sounds. The tool provides two basic analysis aids: the extraction of a single time-frequency parameter on the press of the left mouse button and the extraction of the relative time of a feature from two time reference points (usually the R and T waves of the ECG) on the press of the middle mouse button. This automated extraction of time and frequency information from the time-frequency results facilitates clear and precise extraction of features which
would previously have been extracted manually. All results shown in this work have been displayed and analysed using this tool.

Figure 4.23: X-Windows 'inspect' tool used for the display and analysis of time-frequency distributions.

4.6 Comparative Analysis of Time-Frequency Techniques

This section describes the work carried out to investigate the relative performance of the four time-frequency transforms described above. By applying each technique to two simple sinusoid test signals an initial feel for their resolution capabilities and individual limitations is presented. A conclusive analysis of the suitability of each algorithm to the analysis of heart sounds is presented by applying each technique to a synthetic signal based on real heart sound data.
4.6.1 Time-Frequency analysis of a three sinusoid step signal

To compare the performance of the various time-frequency transforms it is important to apply them to a non-stationary signal. The three sinusoid step signal used in this section exhibits non-stationary properties yet remains relatively simple. At this initial stage, test signal simplicity is desirable to allow a clear definition of expected analysis results. The time-domain and frequency-domain plots of the three sinusoid step signal are shown in Figures 4.24a and 4.24b respectively.

Figure 4.24: Three sinusoid step signal: (a) time-domain representation; (b) frequency-domain representation.
The signal is a three tone frequency stepped signal of length 512 points sampled at 512 Hz. Full parameters of the signal are:

\[ s(n) = 0 \text{ for } 32 > n > 480 \]

\[ f_1 = 16 \text{ Hz for } 32 \leq n \leq 192 \]

\[ f_2 = 72 \text{ Hz for } 193 \leq n \leq 319 \]

\[ f_3 = 128 \text{ Hz for } 320 \leq n \leq 480 \]

Where \( f_1, f_2 \) and \( f_3 \) are the three consecutive frequency values taken by the signal \( s(n) \).

Figure 4.25 shows the results of applying the WT (using a Morlet wavelet), the STFT (using a 64-point window), the WD and the CWD to the three sinusoid step signal. It should be noted that the WT applied in this case is the discrete CWT as defined in (4.10). This implementation of the transform has been preferred to the DWT form of the WT because of the clear visual appearance of the result. The results in Figure 4.25 suggest that the WD is the best technique for separating and localising the signal components. This superior performance is, however, offset by the problem of spurious cross-terms generated by the WD where signal components do not exist. In contrast, the CWD result has no visible cross-terms and has only slightly inferior resolution characteristics. The results generated using the WT and STFT are both free of cross-terms but exhibit inferior resolution characteristics. From these results it would appear that in a situation where the appearance of cross-terms is undesirable (as in the case of heart sound analysis) the CWD is most suited as a time-frequency analysis tool. However, as is pointed out by Jones and Parks [109], the CWD performs particularly well when applied to multicomponent signals in which the components have constant frequencies. For signals with significant frequency modulation the resolution of the CWD degrades. The next section addresses this problem with the application of the transform set to a signal with modulated frequency characteristics.
Chapter 4: Time-Frequency Spectral Estimation

Figure 4.25: Application of four time-frequency techniques to the three sinusoid step signal:
(a) WT; (b) STFT; (c) WD; (d) CWD.

4.6.2 Time-frequency analysis of a dual component LFM test signal

In contrast to the constant frequency step signal used in the previous section, the test signal in this section exhibits extensive frequency variation with time. This signal was constructed as a linear combination of two LFM signals of length 512 points and sampled at 512 Hz. Parameters of these two constituent LFM signals are:-
Chapter 4: Time-Frequency Spectral Estimation

\[
\begin{align*}
    s_a(n) &= 0 \text{ for } 32 \leq n \leq 480 \\
    f_a_1(f) &= 10 \text{ Hz for } n = 32 \\
    f_a_2(f) &= 180 \text{ Hz for } n = 480 \\
    s_b(n) &= 0 \text{ for } 32 \leq n \leq 480 \\
    f_b_1(f) &= 60 \text{ Hz for } n = 32 \\
    f_b_2(f) &= 230 \text{ Hz for } n = 480
\end{align*}
\]

Where \( f_{a_1} \) and \( f_{a_2} \) are the lower and upper frequency limits of the first LFM signal \( s_a \). Similarly, \( f_{b_1} \) and \( f_{b_2} \) are the lower and upper frequency limits of the second LFM signal \( s_b \). Figure 4.26 shows the time-domain and frequency-domain representation of the signal. Figure 4.27 shows the results of applying the WT, STFT, WD and the CWD, in the same format as the last example, to this dual LFM test signal.
Figure 4.26: Dual LFM test signal: (a) low frequency LFM signal $s_a(t)$; (b) high frequency LFM signal $s_b(t)$; (c) combined LFM signal; (d) spectrum of the combined LFM signal.
Figure 4.27: Application of four time-frequency techniques to the dual LFM test signal: (a) WT; (b) STFT; (c) WD; (d) CWD.

Again the results show that the WD is the superior distribution for separating and localising signal components, but again it suffers from significant cross-term interference. Performance of the CWD when applied to this frequency varying signal is significantly poorer than its performance when applied to the fixed frequency component signal. In fact, it has been shown analytically [109] that the CWD, when applied to two frequency modulated signals, can exhibit resolution characteristics inferior to the STFT technique.
Chapter 4 : Time-Frequency Spectral Estimation

The WT, in both examples, shows very poor resolution characteristics, but as outlined in section 4.2 the WT is suited to the analysis of signals containing short high-frequency and long low-frequency components. Neither of the analysed test signals exhibit this behaviour and hence, the poor performance of the WT is not unexpected.

The results in this section and section 4.5.1 all point to a general conclusion that performance of the four time-frequency techniques is very much dependent upon the format of the signal being analysed. Hence, for a true comparative analysis of the performance of the four techniques they must be applied to a signal with characteristics that match those of the signal to be investigated.

4.6.3 Time-frequency analysis of a modelled heart sound

In recent work [5][76] a number of time-frequency methods have been applied to typical heart sound data. Both pieces of work conclude that time-frequency is a technique that shows definite promise as a descriptor of heart sounds. In [76] it is also concluded that it is not possible to make a decision as to the most suitable time-frequency analysis technique due to a poor understanding of the typical heart sounds being analysed. In this section, this problem is overcome by generating a model heart sound with known parameters that correlates very highly with real heart sound data. It is envisaged that with a signal that is very close to the real heart sound data it will be possible to analyse the true performance of the four transforms for this application. Also, with an a-priori knowledge of the test signal characteristics it will be possible to perform a quantitative analysis of the relative performance of the four transforms. A modelled heart sound was generated using the following procedure:-

- Extract a typical first sound from the recordings taken from a subject with normal native valves.
- Model the extracted sound using Sava's modified version of Prony's method for spectral estimation [48]. The resultant model is a linear combination of a known number of decaying sinusoids with known amplitude, frequency, damping factor and phase. The majority of events in the cardiac sound cycle are generally believed to be produced by sudden changes in the position or state of cardiovascular components. This "explosive" origin of sounds results in sounds that have high initial intensity components followed by gradual intensity decay, hence, modelling as a combination of decaying sinusoids is expected to produce a consistent set of signal parameters.
This resultant model with known components is windowed using a Hanning window and padded with zeros at each end to prevent end effect appearing in the time-frequency plots.

Figure 4.28 shows the original sound and modelled sound generated from an extracted first sound recorded from a subject with normal native valves. Given in Table 4.2 are the parameters pertaining to the modelled signal.

![Graph showing comparison between real and modelled signals](image)

**Figure 4.28:** *Time-domain plot of the real first sound recorded from a subject with normal native valves and the equivalent modelled sound.*

<table>
<thead>
<tr>
<th>Amplitude</th>
<th>Frequency</th>
<th>Damping</th>
<th>Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.72</td>
<td>39.54</td>
<td>-0.064</td>
<td>6.52</td>
</tr>
<tr>
<td>0.30</td>
<td>59.49</td>
<td>-0.033</td>
<td>2.18</td>
</tr>
<tr>
<td>0.41</td>
<td>19.30</td>
<td>-0.035</td>
<td>3.77</td>
</tr>
<tr>
<td>0.27</td>
<td>28.28</td>
<td>-0.022</td>
<td>5.19</td>
</tr>
</tbody>
</table>

**Table 4.2:** *Model parameters for the signal generated from the first sound recorded from a subject with normal native valves*
It can be seen from the time-domain plot and from the high value of correlation coefficient that the modelled and original signal are almost identical. This result supports the initial expectation that Prony's model using a combination of damped sinusoids is suited to parameterisation of cardiovascular sounds. Figure 4.29 shows the result of applying the WT, WD, CWD and the STFT, with two alternative window lengths, to the modelled normal first sound with known parameters. The STFT results presented in sections 4.6.1 and 4.6.2 were generated using a constant 64-point (125 ms) window length. Work has been performed into the optimum window choice for heart sound data [74] but results were very signal dependent. In a signal with varying components (i.e. a first heart sound, a second heart sound, a murmur and possibly a pathological click), choice of a window to match all of these component sounds is impossible. Hence, the analysis in figure 4.29 shows results for the STFT using a 32-point (64 ms) and 64-point (128 ms) window length. The frequency resolution of the STFT in these two cases will be approximately 60 Hz and 30 Hz respectively. Considering these resolutions with respect to the frequency values shown in Table 4.2 it is expected that both window lengths will not successfully resolve any more than one or two of the true signal frequency components. On consideration of frequency components alone it can be seen that the components shown in Table 4.2 could be resolved if a longer window length were used in the STFT but it must be remembered that such high frequency resolution would be at the detriment of time resolution. As is shown later in this section the two window lengths used provide a clear trade-off between time and frequency resolution for the modelled heart sound being analysed.

From initial inspection of the results in Figure 4.29 it would seem that the WD and CWD produce almost identical plots both of which are superior, in resolution, to the STFT and WT plots. To qualify this superiority a manual inspection and extraction of significant frequency components for each plot was performed. The results of the analysis of each plot are displayed in Table 4.3.
<table>
<thead>
<tr>
<th>Analysis method</th>
<th>Significant frequency components, Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelet</td>
<td>19, 31, 44</td>
</tr>
<tr>
<td>Wigner</td>
<td>20, 30, 40, 60, 50</td>
</tr>
<tr>
<td>Choi-Williams</td>
<td>20, 30, 40, 60, 50</td>
</tr>
<tr>
<td>32 point STFT</td>
<td>28, 51</td>
</tr>
<tr>
<td>64 point STFT</td>
<td>27, 61</td>
</tr>
</tbody>
</table>

Table 4.3: Results of manual extraction of features from time-frequency plots of the modelled normal first sound
Figure 4.29: Application of four time-frequency techniques to the modelled sound test signal: (a) WT; (b) WD; (c) CWD; (d) 32 point STFT; (e) 64 point STFT.
By comparing these results with the model parameters (Table 4.2) it can be seen that the CWD and WD have picked up all the model frequency components and an extra component at 50 Hz. In the result of (4.21) it was noted that analysis of a simple two frequency \( (f_1, f_2) \) signal using the WD resulted in undesirable cross-terms at the frequency \( \frac{1}{2}(f_1 + f_2) \). Considering this result it is reasonable to suggest that the extra 50 Hz term in the WD and CWD results is a cross-term generated by the interference of the 60 Hz and 40 Hz model components. Comparing the results generated using the STFT it can be seen that using the lower frequency resolution 32-point window none of the model parameters have been correctly identified. Decreasing time resolution and increasing frequency resolution by using a 64-point window results in the the 60 Hz component of the model signal being identified. The WT has performed better than the STFT by resolving successfully two model components, suggesting that the varying window length format of the WT provides an advantage over the fixed window STFT format.

### 4.7 Summary

From the discussion and results in this chapter it can be seen that the WD is the transform that provides the highest time and frequency resolution. This very high resolution performance is unfortunately offset by the inherent cross-term problem that the WD exhibits. The plots of CWD and WD shown in Figure 4.29 are for a single isolated sound and appear almost identical. When a whole cardiac cycle is transformed an extensive mid-cycle interference term, due to interaction of the first and second sound, is generated in the WD plot. The appearance of this interference renders the WD ineffective as a time frequency descriptor of the heart sound cycle signal.

Comparing the results in Figure 4.29 and Table 4.3 it can be seen that the CWD outperforms both the WT and STFT. Results also suggest that the WT performs better than the STFT. Results also suggest that all four techniques have limitations. As already mentioned the WD suffers greatly from the appearance of cross-terms. The WT, due to its multiresolution format, provides a signal representation that must be analysed with these multiresolution characteristics in mind. The WT provides more of a signal representation than a true time-frequency signal description. The obvious limitation of the STFT technique is its poor resolution, but it also suffers from application problems such as choice of window length. Although it can be clearly seen that the CWD used in these examples exhibits less cross-term
components than the WD, it must be remembered that the CWD, as implemented here, is an RID and, as the name suggests, still contains cross-terms. As seen in section 4.6.2, under specific signal conditions these cross-terms can become significant. Hence, interpretation of the results must be made in light of the possibility of cross-term appearance. In past work [5] the CWD has been applied to the analysis of a number of typical first heart sounds to provide a description of frequency dynamics. In section 4.6.3, rather than applying the time-frequency transforms to typical data, a more rigorous approach of estimating the signal parameters prior to analysis was performed. With this more rigorous analysis it was possible to analyse the performance of the transforms quantitatively. Results using the CWD suggested that although a RID it did still exhibit cross-term problems. This new evidence brings into question the results presented in [5] where the possibility of cross-terms within the first heart sound structure was not considered. These results suggest that if the CWD or a similar member of Cohen’s class of distributions is to be used for the detailed description of heart sound frequency dynamics, further investigation of the occurrence and significance of cross-terms should be made.

In light of the results in this chapter it can be concluded that, for the analysis of heart sounds, the CWD exhibits a definite advantage over the STFT, WT and the WD. For this reason, results presented for the time-frequency analysis of heart sounds after this point in the thesis, unless otherwise specified, have been generated with the CWD implemented as described in section 4.4.

The comparison of the four techniques in this chapter has been based upon the requirements of clear visual signal description, integrity of description and high resolution. These are all necessary requirements if the time-frequency techniques are to be used for the visualisation of the recorded heart sounds and subsequent extraction of morphological features (features with clear physical meaning). As already seen, the relative performances of the transforms are very much dependent upon the signal being analysed and will equally be dependent upon the application and analysis requirements. If the application requirements change it is quite possible that the CWD will no longer be the best transform with which to analyse the signal. Later in the text in section 6.3 (Optimised Feature Classification) such a situation arises and an argument is made for the DWT rather than the CWD being the most suitable analysis tool.
Chapter 5

Data Analysis

In the previous chapter the CWD was identified as the most suitable tool for the time-frequency description of recorded heart sounds. This chapter describes the results generated from the application of the CWD to the 100 subject data sets described in section 2.3. The application of the CWD to the native and bioprosthetic valve data in both normal and abnormal cases was investigated in an attempt to demonstrate the diagnostic potential of time-frequency in determining valve condition. Comparing results of time-frequency transforms of various valve types (i.e. native, bioprosthetic and mechanical) it is envisaged that it will be possible to detail the effect that valve type has upon the time-frequency signature of the cardiac sound cycle. Analysing the whole set of results from contrasting valves and contrasting pathological conditions will, it is hoped, provide an insight into the origin of sounds. The origin of heart sounds is a subject that is still not fully understood and any insight is a valuable contribution to improving the understanding of development, detection and treatment of cardiac pathological conditions.

5.1 Time-Domain Sound Pre-Processing and Selection

Previous researchers using classical spectral estimation [72][47] and time-frequency techniques [5] have performed signal preconditioning by averaging a number of sounds to produce a typical sound. This time-domain preprocessing has the advantage that noise with a random format is reduced whilst signal components constructively interfere and an improvement in signal-to-noise ratio is often produced. In this work, unlike these previous researchers, an analysis is being performed on whole cardiac cycles rather than individual
segmented sounds. The whole cardiac cycle, in the case of malfunctioning valves, very often contains extra pathological sounds as well as the usual first and second sounds. These extra sounds can be caused by the turbulent flow of blood over a damaged or ineffective valve. This generation process, because of its turbulent origin, results in sounds that have low correlation from beat to beat, hence, beat to beat averaging often destroys these important sounds. Another problem presented by signal time-domain averaging is the precise matching of individual cycles prior to averaging. Although individual first and second sounds are very repeatable, in fact consecutive discrete sounds can be almost identical, their separation and the duration of a cycle varies extensively. Typically beat separation time can, in a 15 second recording, vary by as much as 20%. For these reasons it was decided that no time-domain averaging was to be performed on the data and an analysis was to be performed on an extracted typical sound. This subjective extraction process presents a problem of result reproducibility hence a formal procedure for the extraction of typical sounds was formulated and performed for all the results presented in this work.

5.1.1 Typical sound extraction

Sound extraction was performed with two important requirements in mind: representative results and clear results. To ensure representative results, the whole recording was first viewed to identify any sounds that were clearly affected by recording inconsistencies such as patient movement or external ambient sound. These sounds were not considered for analysis. To ensure clear results the sounds were vetted for consistent amplitude, cycles were extracted that showed typical relative first and second sound intensities. This conservation of relative sound intensity is desirable because in the past this quantity has been used, both in automatic systems and auscultation, as a diagnostic parameter. In pathological cases it was very often observed that the PCG signal could undergo what could be described as "collapse". In these instances the signal clearly generated an unusual sound. In this case the sound and its adjacent sounds were disregarded and typical sound selection was made from the remaining examples. This collapse of the cycle and its causes obviously present diagnostic information but the work in this thesis concentrates on the pathological information presented in a single cardiac cycle, hence, this diagnostic information was not taken into consideration and was discarded. This identification process is biased towards the selection of the most normal sound in the recording and hence, analysis is performed after first discarding diagnostic information. Although this sound selection process increases the difficulty of the diagnostic task it is felt that if sound analysis were based on favourable data results would be far from representative.
Before a typical sound cycle could be extracted it was necessary to make a decision as to the length of data to be extracted. The analysed sounds were recorded at a sampling rate of 5 kHz. To maintain this sampling frequency and to extract a whole cardiac cycle of sounds it would be necessary to extract at least 5000 samples (on average the heart in a normal resting subject will beat 70 times a minute). The 5 kHz sampling rate allows the recorded sounds to be analysed up to 2500 Hz. Analysis to this high frequency point is unnecessary, the significant sounds generated by biological valves are always below 250 Hz and in the majority of cases below 150 Hz. For this reason the analysis of biological valve sounds is only performed up to 250 Hz. In the case of metallic valves the upper frequency limit to significant sounds is documented as high at 10 kHz. To accommodate these high frequency characteristics the metallic sounds are analysed in a number of stages. The first stage is a low frequency analysis up to 250 Hz. Following this, areas of high frequency were identified and analysed up to the full Nyquist limit of 2500 Hz. Analysis was not extended above this point because initial investigation of the data using both time-frequency and classical spectral estimation techniques indicated no significant sound components above this point. Prior to the low frequency analysis of sounds the signal was subsampled by a factor of 10. The resulting signal stored at a 500 Hz sampling rate was inspected for typical sound cycles over a 1024 ms / 512 sample period.

Although all subjects were in a state of relaxation during recording sessions, there was extensive variation in heart rates from patient to patient. Variation in the heart rate of recorded subjects presents a number of practical problems when analysis is performed over a fixed 512-point window. The first problem is in the case where a subject’s heart rate is well in excess of the normal 70 beats per minute. In these situations the analysis window may contain more than one first and second sound. Without knowing the full context of the displayed sounds there is a possibility that these extra sounds may be misinterpreted as pathological sounds. An additional problem presented by multiple cycles being included in the analysis is the generation of cross-terms between cycles. As concluded in chapter 4 the CWD as an RID does still exhibit an amount of signal interference and the inclusion of repeated sounds in an analysis will also result in the inclusion of unnecessary interference terms. To prevent the inclusion of multiple cycles in the extracted typical sound cycles, the raw 512 sample sections are windowed prior to time-frequency analysis. The window function applied is:

$$ s'(n) = s(n) \times \text{win}(n) \quad 0 \leq n \leq M $$

(5.1)

where $s'(n)$ is the new signal generated after multiplication of the original signal $s(n)$ by the window function $\text{win}(n)$ and $M$ is the number of signal points (in this case 512).
description of the method used to extract the beat length for each subject's ECG recording is detailed in section 5.1.2. The window function \( \text{win}(n) \) is defined as:

\[
\text{win}(n) = 0.5 - \left[ 0.5 \cos \left( \frac{2\pi n}{N - 1} \right) \right] \quad 0 \leq n \leq N/2
\]

\[
\text{win}(n) = 1 \quad N/2 \leq n \leq M - N/2
\]

\[
\text{win}(n) = 0.5 - \left[ 0.5 \cos \left( \frac{2\pi (n - M + N)}{N - 1} \right) \right] \quad M - N/2 \leq n \leq M
\]

Where \( N \) is defined as the beat length of recorded sounds subtracted from the signal length \( M \). In effect the window function is a Hanning window with an extended region at its centre, over which the integrity of the signal is maintained. This maintained region is equal in length to the known beat period of the recorded sound data. Using this pre-windowing technique we can guarantee to maintain the integrity of a full cardiac cycle of information while preventing the inclusion of any significant amount of information from adjacent sound cycles.

### 5.1.2 Extraction of time information from the ECG

The recording of the PCG signals were accompanied by a synchronous recording of the subject's ECG lead II signal. The ECG provides a very clear simple signal that can be used to extract concise timing information about the recording. Figure 5.1 shows a schematic representation of the relationship between ECG and PCG signal.

![Figure 5.1: Timing relationship between ECG and PCG in a normal subject.](image)
Included on the sketch are the five recognised P, Q, R, S, T deflections of the ECG signal and simplified representations of the PCG first sound (1) and second sound (2). It can be seen that the first sound and QRS complex of the two signals are coincidental and the second sound occurs at the end of the T wave of the ECG. With these relationships in mind it was decided that extraction of two features from the ECG would allow complete time referencing of the PCG signal. These two points were the maximum point of R wave deflection and maximum point of T wave deflection. Extracting these two time-domain features of the ECG provides a reference to the first and second heart sound respectively. Another important feature of the recording that can be clearly extracted from the ECG is the pulse rate or the beat to beat separation. Due to significant variations in pulse rate over a single recording it was decided that a localised pulse rate in the region of the sounds of interest should be extracted. This was accomplished by measuring the time between four successive R wave maxima. The average time of these four adjacent beats was recorded as the patient’s beat to beat interval.

By extracting this ECG information and extracting a representative and clear PCG cardiac cycle it was felt that the data was available for the extraction of any available diagnostic features.

5.2 Time-Frequency Analysis of Native Valve Sound Data

Past work investigating the usefulness of various spectral estimation techniques to the analysis of heart sounds has, in the majority of cases, concentrated on sounds generated by implanted prosthetic valves. Analysis of native valves has also been performed to identify spectral normality in healthy subjects. An application that has surprisingly been almost ignored is the use of spectral estimation for the diagnosis of normal and abnormal native valve sounds. This section illustrates a number of typical results for the analysis of native heart valve sounds and addresses three issues: the suitability of time-frequency to the task of native valve diagnosis, the identification of the effect recording position has on distribution result and the effect subject variability has on distribution results. Figures 5.2, 5.3 and 5.4 show six contrasting time-frequency distributions for normal, abnormal mitral, abnormal aortic and abnormal double native valve subjects. Included on each time-frequency plot are two vertical lines indicating the position of the R and T wave of the synchronous ECG trace.
Figure 5.2: Time-Frequency distributions of native valve sounds: (a) abnormal aortic valve; (b) normal native valves.
Figure 5.3: Time-Frequency distributions of native valve sounds: (a) abnormal mitral valve; (b) normal native valves.
Figure 5.4: Time-Frequency distributions of native valve sounds: (a) double aortic and mitral abnormal valves; (b) normal native valves.
One of the objectives of this work has been to prove the usefulness of time-frequency descriptions for the identification of abnormal valves. Figures 5.2, 5.3 and 5.4 give an initial feel for the problem. Figure 5.2 shows the time-frequency distributions of recordings taken from an abnormal aortic native valve subject and a normal native valve subject respectively. Both recordings were made with the microphone placed in the aortic position on the patient's chest. It is very easy to see the contrast between the normal and abnormal case due to the appearance of a substantial mid-systolic murmur. From this example, it would seem that valve dysfunction results in the time-frequency distribution taking on a definite appearance of abnormality. As described in chapter 2 the systolic murmur in Figure 5.2, assuming mitral valve normality, is most probably due to turbulent blood flow through a stenosed aortic valve. Turbulence, as an explanation of the generation of these murmurs, is supported by the murmur's high frequency and random appearance in the time-frequency domain. The structure of murmurs in the distributions generated for the remainder of the population resemble that shown in Figure 5.2a. Hence, results consistently support the hypothesis that turbulent blood flow is the dominant mechanism behind the generation of murmurs.

Comparing the time-frequency distributions in Figures 5.2a and 5.2b with respect to their relationship with the two time reference points, it is obvious that there are timing differences. The two distributions show considerably different time for the separation of the second sound and the T wave of the ECG. It would seem from the consistent separation of the first and second sounds, in all the examples in Figures 5.2, 5.3 and 5.4, that in the dysfunctioning case of Figure 5.2a the T wave of the ECG is delayed rather than the second sound being early. This delay in the T wave of the ECG may be due to the maintenance of tension in the ventricle due to the high ventricle pressure required to force blood through the dysfunctioning valve.

Considering frequency characteristics, it can be seen that in both cases of Figures 5.2a and b the first and second sounds are concentrated about 30 Hz with the first sound being of longer duration and the second sound reaching higher frequencies. This basic appearance of the two major sounds is consistently shown throughout the the normal examples in Figures 5.2b, 5.3b and 5.4b. An exception to this occurs in dysfunction when a sound may be missing or swamped by a pathological murmur.

In contrast to the aortic case, the dysfunctioning mitral case and the normal subject recorded in the mitral position (Figure 5.3) appear to be very similar. In fact, comparing the dysfunctioning mitral case (Figure 5.3a) to all three normal cases (Figures 5.2b, 5.3b, 5.4b) it can be
seen that the dysfunctioning mitral case is almost indistinguishable from normal subjects. The only possible difference between Figure 5.3 and the normal recordings is a quiet low frequency sound after the second sound between 700 ms and 800 ms. From the description in chapter 2 it may be concluded that this sound is either a third heart sound, a fourth heart sound or an early diastolic murmur. There also seems to be a delay in the T wave of the ECG but it is less pronounced than in the aortic case. This similarity with the normal population is demonstrated in the majority of the dysfunctioning native mitral population. A few exceptions to this appearance of normality show a systolic murmur possibly caused by blood leaking through the closed mitral valve. From this initial look at the time-frequency pattern of mitral valve dysfunction it would seem that identification of dysfunctioning native mitral valves poses a particularly difficult problem.

The third and final example of contrasting dysfunctioning and normal subjects (Figure 5.4) are sounds recorded in the aortic position from a dysfunctioning double valve subject and again a normal subject. The term dysfunctioning double valve refers to a subject diagnosed as suffering from both aortic and mitral valve disease. These results show a definite contrast between the normal and dysfunctioning case. In the dysfunctioning case the first sound seems to have gained a significant high frequency component or alternatively there is a very early systolic murmur. In addition, the second sound seems to have risen in frequency and discarded the majority of its low frequency components. Although there is no indication of any extra pathological sounds it is clear that there has been definite modulation of the first and second sound components.

Figures 5.2, 5.3 and 5.4 all show significant sound components below 50 Hz and, as detailed in chapter 3, filtering does occur below this frequency. The filter used is a low order filter with a gradual roll-off, hence, the filter removes very low frequency noise signals but has only a small effect upon the major heart sound components just below 50 Hz.

Results in Figures 5.2, 5.3 and 5.4 show that time-frequency offers definite potential as a tool for the identification of abnormal sounds recorded from dysfunctioning aortic and double valve subjects. In contrast, results for dysfunctioning mitral valve subjects would suggest that time-frequency offers little potential for diagnosis in the mitral case. In general these typical results suggest that the identification of dysfunctioning subjects should be performed using both time and frequency features with the significance of features being dependent upon the condition of the subject being analysed.
Chapter 5: Data Analysis

As described in section 5.1.1 the signals and results shown in Figures 5.2, 5.3 and 5.4 are windowed versions of the true recordings. This windowing procedure was justified as a means of reducing misinterpretation of multiple heart sound cycles and removing undesirable inter-beat interference. With respect to these objectives it can be seen from Figures 5.2, 5.3 and 5.4 that the windowing procedure has been highly successful. In contrast to this success, the windowing procedure does have an undesirable side effect. Windowing the signal about a single beat involves discarding information about inter-beat dependencies and relationships. However, no attempt is made in this work to investigate this aspect of the results.

For practical reasons results presented in this section are given as typical examples of the whole data population. In an attempt to present an outline of the results for the complete population nine features describing the general appearance of the time-frequency distributions were extracted for each data subject. The nine features used are given in Table 5.1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N12</td>
<td>How many of the two principle sounds S1, S2 are present?</td>
</tr>
<tr>
<td>2</td>
<td>NS</td>
<td>How many extra sounds are present, not including S1 and S2?</td>
</tr>
<tr>
<td>3</td>
<td>f1</td>
<td>Frequency of S1 maximum intensity (Hz)</td>
</tr>
<tr>
<td>4</td>
<td>f2</td>
<td>Frequency of S2 maximum intensity (Hz)</td>
</tr>
<tr>
<td>5</td>
<td>f1max</td>
<td>Highest S1 frequency (Hz)</td>
</tr>
<tr>
<td>6</td>
<td>f2max</td>
<td>Highest S2 frequency (Hz)</td>
</tr>
<tr>
<td>7</td>
<td>S1-S2</td>
<td>Time between features f1 and f2 (ms)</td>
</tr>
<tr>
<td>8</td>
<td>R-S1</td>
<td>Time from ECG R wave to feature f1 (ms)</td>
</tr>
<tr>
<td>9</td>
<td>T-S2</td>
<td>Time from ECG T wave to feature f2 (ms)</td>
</tr>
</tbody>
</table>

Table 5.1: Nine descriptive features extracted from sound time-frequency distributions

Features 1 and 2 describe the consistency of the two major cardiac sounds (the first and second sounds) and the appearance of extra pathological sounds. Features 3-6 describe the frequency characteristics of the first and second sound. Features 7-9 describe the timing of the first and second sounds with respect to each other and to the two ECG features. Figure 5.5 shows an illustration of how, using the tool described in section 4.5, these nine features are extracted. In this example it can be clearly seen that there is both a first and second heart sound, hence, feature N12 takes the value 2. The distribution in Figure 5.5 contains a significant mid-systolic murmur and hence, with this extra component the feature NS takes a value of 1. It should also be noted that in the example of Figure 5.5 the feature T-S2 has a negative
value. As can be seen in Figures 5.2, 5.3 and 5.4 this feature will actually be positive in the majority of cases. The distribution of the nine features in Table 5.1 for the four groups: normal aortic, dysfunctioning aortic, normal mitral, dysfunctioning mitral are illustrated in Table 5.2 in terms of a feature mean and standard deviation. Results for N12 and NS are presented in the form 10/2, indicating that 10 of the group members have 2 sounds.

**Figure 5.5:** Illustration of the extraction of the time-frequency features in Table 5.1 from a typical result.

<table>
<thead>
<tr>
<th>Group</th>
<th>N12</th>
<th>NS</th>
<th>f1</th>
<th>f2</th>
<th>f1max</th>
<th>f2max</th>
<th>S1-S2</th>
<th>R-S1</th>
<th>T-S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm aortic</td>
<td>10/2</td>
<td>10/0</td>
<td>36.4</td>
<td>35.9</td>
<td>87.9</td>
<td>128.0</td>
<td>354.9</td>
<td>43.4</td>
<td>127.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.2</td>
<td>8.7</td>
<td>15.1</td>
<td>35.9</td>
<td>22.5</td>
<td>16.9</td>
<td>17.1</td>
</tr>
<tr>
<td>dysfunc aortic</td>
<td>1/0, 2/1</td>
<td>2/0, 10/1</td>
<td>41.0</td>
<td>52.4</td>
<td>111.3</td>
<td>122.4</td>
<td>340.8</td>
<td>65.8</td>
<td>66.2</td>
</tr>
<tr>
<td>norm mitral</td>
<td>10/2</td>
<td>10/0</td>
<td>12.9</td>
<td>33.9</td>
<td>22.6</td>
<td>41.8</td>
<td>28.6</td>
<td>26.8</td>
<td>64.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12.9</td>
<td>33.9</td>
<td>22.6</td>
<td>41.8</td>
<td>28.6</td>
<td>26.8</td>
<td>64.5</td>
</tr>
<tr>
<td>dysfunc mitral</td>
<td>9/0, 3/1</td>
<td>3/1, 9/2</td>
<td>36.8</td>
<td>30.5</td>
<td>116.5</td>
<td>80.6</td>
<td>326.1</td>
<td>71.2</td>
<td>107.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17.6</td>
<td>18.2</td>
<td>34.2</td>
<td>32.0</td>
<td>41.4</td>
<td>32.2</td>
<td>15.7</td>
</tr>
</tbody>
</table>

**Table 5.2:** Distribution of N12 and NS results together with mean and SD values for the seven T-F parameters for all four native population groups.
In the case of the parameters NS and N12 it is very easy to define normality. Table 5.2 shows normal subjects in both the mitral and aortic cases contain, without exception, only the first and second heart sounds. For the remaining seven parameters, specifying normality is more difficult. An interesting feature in Table 5.2 is the fact that for the aortic case standard deviation in each feature is higher for the abnormal subjects than for the normal subjects. This would suggest that in normal subjects the time and frequency features are concentrated about a normal value. Also, in the dysfunctioning case the time-frequency structure of the cardiac sounds is modified such that these features of interest deviate away from their region of normality. Again, this evidence suggests time-frequency has definite potential as an aid to the diagnosis of dysfunctioning native aortic valves. Diagnosis may be indicated by the appearance of extra pathological sounds or the deviation of the first and second sound time-frequency characteristics from a defined region of normality.

From Table 5.2 it can be seen that in contrast to the aortic case, standard deviation values in the mitral valve case do not support the hypothesis of a region of time-frequency normality for normal valve subjects. For example feature R-S1, in the abnormal case, shows a considerably lower value of standard deviation than in the normal case. In spite of this it can be seen that this feature shows very different mean values between normal and abnormal subjects. These contrasting mean values suggest that although the normal subject feature values are spread they are in fact separated from the abnormal values and again the groups may well be separable.

Results in Figures 5.2, 5.3 and 5.4 demonstrate that time-frequency provides a clear description of the composition of the cardiac sound cycle. The time-frequency nature of the description these plots provide results in an analysis that bears a close resemblance to auscultation techniques. The comparison of the whole population with respect to each other, as shown in Table 5.2, clearly demonstrates that these plots contain diagnostic information. The results in Figures 5.2, 5.3, 5.4 and Table 5.2 highlight, particularly in the mitral case, that identification of dysfunction is a complex problem. By comparing all three typical normal distributions (Figures 5.2b, 5.3b 5.4b) it can also be seen that although there is an amount of consistency, there is a great deal of variability between these members of the same class. Again, this characteristic highlights the complexity of the problem of classifying valve condition via the analysis of surface recorded heart sounds.
5.3 Time-Frequency Analysis of Bioprosthetic Valve Sound Data

A common application of spectral estimation when applied to heart sounds is based around the detection of dysfunction in subjects with implanted bioprosthetic heart valves. As described in earlier chapters the development of a tool for the analysis of bioprosthetic valves is desirable due to the limited duration and inevitability of dysfunction in this type of valve. Presented in Figures 5.6 and 5.7 are typical examples of time-frequency distributions generated from sounds recorded from normal/abnormal subjects with C-E bioprosthetic valves in the aortic and mitral positions respectively. Again, as in the native valve case, it would seem that normality is characterised by the sound cycle exclusively containing first and second sounds components. It would also seem that dysfunction in the case of aortic bioprosthetic valves (Figure 5.6a) is, as in the native case (Figure 5.2a), characterised by the appearance of a definite systolic murmur. Of particular interest in the result of Figure 5.6 is the almost complete absence of a first sound and the dominance of the second heart sound in the dysfunctioning case. The result shown in Figure 5.6 is, as described in [110], an almost typical example of the heart sounds that would be expected for a patient suffering from aortic valve disease and exhibiting severe aortic regurgitation. As described by Dawkins [110] the first sound is quiet due to premature mitral closure and there is clear evidence of an early diastolic murmur. The modification of the first heart sound during aortic valve dysfunction suggests that the first and second sounds can not be simply accredited to the mitral and aortic valves respectively. In fact, this result clearly demonstrates that the appearance and nature of the first and second heart sounds are governed by complex cardiac systemic factors.
Figure 5.6: Time-Frequency distributions of C-E bioprosthetic valve sounds recorded in the aortic position: (a) Abnormal; (b) Normal.
Figure 5.7: Time-Frequency distributions of C-E bioprosthetic valve sounds recorded in the mitral position: (a) Abnormal; (b) Normal.
A feature of interest that appears in both Figure 5.6b and 5.7b is the high frequency nature of the aortic and mitral valve sounds respectively. These high frequency characteristics in the examples in Figures 5.4b and 5.6b suggest that highest first sound frequency and highest second sound frequency may be useful features in determining the condition of bioprosthetic valves. The correlation of these high frequency characteristics with valve condition would seem to suggest that a normal valve closes efficiently, creating a clean crisp sound. This connection of valve condition with the high frequency nature of the recorded sounds supports the possibility of a degree of connection between the first and second sounds and the vibration of mitral and aortic valves respectively. In contrast to the native valve results it can be seen that there is no obvious connection between the timing of the major sounds with respect to the ECG and the condition of the implanted bioprosthetic valve. The only possible explanation for this inconsistency is the different structures of the native and bioprosthetic valves. With different structures it is possible that failure occurs in different forms and hence, the effect of failure on the operation of the heart is also different.

The results in Figures 5.6 and 5.7, although described as typical, are not representative of the whole data set. Figure 5.8 shows the time-frequency distributions of two subjects taken from the normal aortic C-E valve population. It can be seen that both these "normal" valve subjects exhibit characteristics that, because of their similarity with abnormal C-E and native valve subjects, suggest a state of aortic dysfunction. These subjects were individuals who were active participants in an outpatient follow-up program designed to monitor valve integrity after surgery.
Figure 5.8: Time-Frequency distributions of 'normal' C-E bioprosthetic valve sounds recorded in the aortic position.
Results in Figure 5.8 suggest that these subjects had aortic bioprosthetic valves that were in a state of dysfunction that had not, up to the date of recording, been detected by present diagnostic techniques. This result suggests that the time-frequency analysis of recorded sounds offers a viable alternative to present outpatient valve monitoring techniques. As in the native valve case, the nine features in Table 5.1 were extracted from the full bioprosthetic data set to provide a description of the population and point to possible areas of diagnostic potential. The distribution of these nine features for the three groups: normal aortic, dysfunctioning aortic, normal mitral, are illustrated in Table 5.2 in terms of mean and standard deviations for each population.

<table>
<thead>
<tr>
<th>Group</th>
<th>N12</th>
<th>NS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>norm aortic</td>
<td>2/1</td>
<td>9/0</td>
<td>3/1</td>
<td>mean</td>
<td>30.1</td>
<td>55.2</td>
<td>83.1</td>
<td>165.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SD</td>
<td>11.4</td>
<td>27.9</td>
<td>19.5</td>
<td>51.5</td>
</tr>
<tr>
<td>dysfunc aortic</td>
<td>3/1</td>
<td>4/0</td>
<td>5/1</td>
<td>mean</td>
<td>35.0</td>
<td>40.5</td>
<td>105.1</td>
<td>134.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2/2</td>
<td></td>
<td>SD</td>
<td>11.7</td>
<td>7.8</td>
<td>38.7</td>
<td>27.5</td>
</tr>
</tbody>
</table>

Table 5.3: Distribution of N12 and NS results together with mean and SD values for the seven T-F parameters for both CE population groups. Results for N12 and NS are presented in the form 10/2, indicating that 10 of the group members have 2 sounds.

In the native valve case shown in Table 5.2 it was noted that, without exception, the normal valve subjects had only first and second sound components. It can be seen in Table 5.3 that in the majority of cases this exclusive appearance of the first and second sounds also occurs for the bioprosthetic population, but as demonstrated in Figure 5.8 there are exceptions. Considering the frequency characteristics $f_2$ and $f_{2\text{max}}$ in normal bioprosthetic valves it can be seen that average results support the high frequency characteristics shown in Figure 5.6. By comparing first sound frequency characteristics from native and bioprosthetic results, it can be seen that on average very little difference is evident. These results would suggest that aortic valve replacement with a bioprosthetic valve has little, if any, effect upon the characteristics of the first heart sound and causes an increase in the frequency characteristics of the second heart sound.

By comparing the standard deviations for normal and dysfunctioning bioprosthetic valves it can be seen that in the majority of features the normal sounds show the largest variation. This variation in the normal feature values may be due to the class uncertainty of subjects such as
those shown in Figure 5.8.

In the case of bioprosthetic valves it can be seen that the largest feature change from normality to dysfunction is the drop in f2max. This confirms the observations made on the typical examples in Figure 5.6. Finally, by comparing the change from normal to dysfunctioning in both native and bioprosthetic valve cases it can be seen that in six of the seven features a common direction of change is exhibited. This similar change in feature values for native and bioprosthetic valve dysfunction suggests that in both cases dysfunction causes a similar modification to first and second sound time-frequency characteristics.

5.4 Time-Frequency Analysis of Mechanical Valve Sound Data

The material construction of bioprosthetic and native valves is very similar and hence, the results form both normal and abnormal cases show many similarities. In recent years, the number of newly implanted bioprosthetic valves has dropped and the majority of native valve replacements are now being performed with modern mechanical prosthesis. The popularity of these mechanical prosthesis over their traditional biological equivalents is, in the majority of cases, due to their superior durability and hence, the reduced incidence of valve failure. Due to their recent popularity it is important that we investigate here the suitability of time-frequency techniques to the description and analysis of mechanical valve sounds.

To this point, native and bioprosthetic valve sounds have been analysed in the frequency range 0 - 250 Hz and in a number of cases only 0 - 175 Hz of the result has been displayed. Due to the material construction of mechanical valves they have been documented as emitting frequencies up to and above 15 kHz [53]. The data recorded for use in this work has been sampled at 5 kHz and hence, provides a possible analysis up to a 2.5 kHz Nyquist frequency. Analysis of a complete sound cycle using the CWD would require a signal sample length of approximately 5000 points. The analysis of such a large set of samples presents a number of practical problems. In particular, due to the inherent expansion of this one-dimensional data into a two-dimensional distribution, hardware memory requirements become excessive. One possible solution to this problem is a two stage analysis in the form of an initial low frequency analysis (0 - 250 Hz) of the whole cardiac cycle followed by a high frequency analysis (0 - 2500 Hz) of the areas of the cycle that show signs of high frequency activity. Figure 5.9 shows the time-frequency distribution of two normal aortic metallic valves analysed over the low frequency range.
Figure 5.9: Time-Frequency distributions of normal Bjork-Shiley prosthetic valve sounds recorded in the aortic position.
Chapter 5 : Data Analysis

Figure 5.9 quite clearly shows a significant high frequency component in the second heart sound that is not apparent in equivalent native and bioprosthetic examples. It should also be noted that the majority of first and second sound energy is still concentrated in the region of 30 Hz. This result suggests that, as in the case of bioprosthetic valves, replacement of a native aortic valve with a mechanical prosthetic valve has a definite effect on the initial component of the second sound. This correlation of second sound and aortic valve characteristics clearly supports a hypothesis that valve replacement has an effect on the frequency characteristics of the associated sound. The concentration, irrespective of valve type, of first and second sounds about 30 Hz suggests a common origin to the major components of the sounds. Due to the contrast between biological and mechanical valve construction and vibration characteristics it is valid to say that the consistent major component of the first and second sound cannot be generated by valve structural vibrations. Hence, these results bring into question the validity of past works [11][12][13][14] that have attempted to correlate valve structure with the frequency of the major sound components.

It is important to note that the results in Figure 5.9 for normal aortic mechanical valves show, at low frequencies, a high degree of similarity to those given earlier for normal native and bioprosthetic valves. This would suggest that it is possible to clearly define the appearance, in the time-frequency domain, of subject normality irrespective of valve type.

The analysis shown in Figure 5.9, with a frequency range of 0 - 250 Hz, does not adequately cover the frequencies over which metallic valve sounds appear. To allow the analysis of the high frequency characteristics of these valves, areas of high frequency activity must be identified. The identification of high frequency components in the signal is hampered by the dominance of low frequency sound components. The dominance of these low frequency components in the heart sounds of mechanical valve patients contradicts the general appearance of sounds when analysed using auscultation. In subjects with implanted mechanical valves there is a dominant high frequency click associated with valve closure that can often be heard with the naked ear up to a few feet from the patient. The apparent dominance of this high frequency valve vibration is due to the high-pass nature of the ear’s frequency response. In an attempt to highlight the high frequency components of the metallic prosthetic valve sounds the recorded data can be similarly high-pass filtered prior to analysis. The frequency response of the human auditory system is a quantity that varies with sound intensity and frequency [115] and, as such, it is very difficult quantity to model. In an attempt to produce an approximation to the ear’s response over the 20 - 2500 Hz region of interest, a very simple definition of a high-pass filter was proposed. Figure 5.10 shows an illustration of the
specification used to design a filter to mimic the frequency response of the human auditory system.

Figure 5.10: Filter specification for modelling the human auditory response to heart sound frequency information.

A close approximation to the specification in Figure 5.10 was obtained by implementing a second order high-pass Butterworth filter with a cut-off frequency at 300 Hz. Figure 5.11 illustrates the effect of this high-pass filter when applied to a complete sound cycle recorded from a subject with a normal aortic metallic prosthetic valve. It can be seen from Figure 5.11 that together with the initial second sound high frequency component there are definite high frequency components that occur within the region of the first heart sound. Considering that these high frequency components can only have been produced by the metallic structure of the aortic prosthetic valve [53] it is reasonable to assume that they are generated as the valve opens. Similar opening sounds have been reported for other prosthetic valve types [116]. Figure 5.12 shows the time-frequency distributions of these high frequency opening and closing prosthetic sounds.
Figure 5.11: Application of a high-pass filter to sounds recorded from a subject with a normal aortic metallic prosthetic heart valve: (a) the original signal; (b) complete filtered signal; (c) amplitude scaled metallic valve opening sounds; (d) metallic valve closing sounds.
Figure 5.12: Time-Frequency distributions of metallic valve sounds: (a) opening sound shown in Figure 5.11c; (b) closing sound shown in Figure 5.11d.
The time-frequency distribution of the opening metallic sound in Figure 5.12a shows two definite, independent components with maximum energy at approximately 250 Hz. By inspecting the distribution closely it can be seen that the maximum intensity of the second component of the opening sounds is slightly higher in frequency than the first component. It should also be noted that there is, even after high-pass filtering, still an amount of low frequency information that is quite probably due to the coincident first heart sound. It can be seen that all three components of the opening and closing sounds have frequency components well in excess of the previous 250 Hz analysis limit. The maximum frequency reached by the components in Figure 5.12 is approximately 1 kHz which is well below the 15 kHz maximum reported by previous researchers [53]. There are two possible explanations for these missing high frequencies. Firstly, their very small relative intensity renders them insignificant in comparison to the low frequency, higher intensity sound components. Secondly, previous works reporting high frequency metallic valve sounds have been performed in in-vitro studies where the lung-thorax system has not been considered. In this in-vivo study the lung-thorax system, due to its low-pass frequency response [117], causes attenuation of the high frequency metallic sounds again rendering them insignificant in comparison to the higher intensity low frequency components.

The results in Figure 5.12a suggest that during valve opening two individual events occur. The subject analysed in Figure 5.12 had a normally functioning Bjork-Shiley tilting disc valve implanted in the aortic position. Due to the construction of this type of valve, opening occurs in two distinct stages. Firstly, the valve disc tilts to an open position and secondly, the disc slides to an improved flow position [118]. It is quite possible that the dual structure of the opening sound is a result of this two stage valve opening. Past PCG investigations of prosthetic valve sounds [116] have also shown multiple valve clicks for other valve types. The correlation of these recorded sounds with the mechanisms of valve operation provides diagnostic potential. In the event of valve failure it is quite possible that the structure of this open and slide signature will be modified.

It can be seen from Figure 5.12b that valve closure generates sound components up to a frequency of 800 Hz. Such high frequency components are undoubtedly generated by vibration of the metallic valve structure. The appearance of these valvular vibration sound components supports the hypothesis that aortic valve vibration contributes to the recorded second heart sound. Due to the short duration of these high frequency components, valve structure vibration can only be connected with a very limited portion of the total second sound energy.
Chapter 5: Data Analysis

As in the native and bioprosthetic valve cases a number of descriptive features were extracted from the time-frequency distributions of the whole metallic valve population. Table 5.4 illustrates the distributions of these nine features. For consistency with other valve results the complete analysis of the metallic valve population was performed for complete sound cycles over the low frequency (50 - 250 Hz) range.

<table>
<thead>
<tr>
<th>Group</th>
<th>N12</th>
<th>NS</th>
<th>f1</th>
<th>f2</th>
<th>f1max</th>
<th>f2max</th>
<th>SI-S2</th>
<th>R-S1</th>
<th>T-S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm aortic</td>
<td>14/2</td>
<td>130, 1/1</td>
<td>mean</td>
<td>31.9</td>
<td>53.1</td>
<td>72.1</td>
<td>203.8</td>
<td>306.0</td>
<td>91.2</td>
</tr>
<tr>
<td>SD</td>
<td>8.0</td>
<td>24.5</td>
<td>13.5</td>
<td>30.3</td>
<td>32.6</td>
<td>37.2</td>
<td>28.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Distribution of N12 and NS results together with mean and SD values for the seven T-F parameters for subjects with implanted aortic Bjork-Shiley valves. Results for N12 and NS are presented in the form 10/2, indicating that 10 of the group members have 2 sounds.

In the case of subjects with implanted aortic metallic valves, unsurprisingly, it can be seen that maximum second sound frequency is considerably higher than in the native and bioprosthetic valve cases. In the remaining feature cases there is little difference between metallic and biological average values. These results suggest, as demonstrated in the typical examples of Figure 5.9, that aortic valve structure only has an effect on a very short period at the onset of the second sound. Again, it can be seen that the exclusive appearance of the first and second sound is a characteristic of valve normality.

5.5 Time-Frequency Analysis of Pre & Post Operative Valve Sound Data

To this point, the work presented in this chapter has concentrated on either the identification of features that may indicate valve dysfunction or the contribution of valve vibration to the recorded heart sounds. A significant amount of past research [117][4][80] has been directed towards the identification of the effect the lung-thorax system has upon surface recorded heart sounds. Although it is generally accepted that the lung-thorax system has a substantial effect on surface recorded heart sounds its relationship to the frequency and time structure of the sounds has yet to be established.
In an attempt to investigate the influence of the lung-thorax system on the cardiac sound cycle the heart sounds of a number of patients were recorded one day before and four to six days after mechanical heart valve implantation. Comparing these before and after surgery recordings provides information on the influence of both the common lung-thorax system and the contrasting valve types on the structure of the recorded sounds. The origin of sounds and the nature of the mechanisms that effect their transmission can be modelled as the simple structure shown in Figure 5.13.

Figure 5.13: Model of the system of generation and transmission of heart sounds.

In Figure 5.13, $s_{in}$ represents the heart sound source. Due to the uncertainty that surrounds the knowledge of sound origin the proposed model has an input system composed of the whole heart-valve system (i.e. myocardium tissues, adjacent vessels and the contained blood). Added to this original signal, before transmission, is the noise $n_{in}$ generated from internal mechanisms such as respiration and thoracic muscular activity. The quantity $n_{out}$ represents noise added to the signal at the chest surface. Contributions to this external noise will include ambient room noise and electronic instrumentation noise. The resultant recorded signal $s_{out}$ will be a combination of the true heart sound signal and the internal and external noise sources. It is desirable to use a recording procedure that minimises the contribution of the two noise signals $n_{in}$ and $n_{out}$ to the overall signal $s_{out}$. The influence of these noise components is reduced via two methods. Firstly, the recording of data is limited to frequencies over which heart sounds are known to occur, discarding signal components outside these frequencies of interest result in an improved signal-to-noise ratio. The effect of internal noise is reduced by recording the sounds in one of two predefined aortic or mitral positions (see chapter 3). It is widely accepted that these positions are the sites over which heart sounds appear loudest and as such present the best signal-to-noise ratio of recorded sounds [1]. Considering these recording procedures it is possible to assume that the interaction of the heart valve system and the lung-thorax system is the primary mechanism governing the
appearance of the surface recorded sounds. Figures 5.14 and 5.15 show the before and after time-frequency distributions for subjects with a dysfunctional aortic and mitral valve respectively. As in previous examples, the position of the R and T waves of the synchronous ECG are indicated by two vertical black lines.

In Figure 5.14 it can be seen that the before surgery distribution (Figure 5.14a) shows definite characteristics of dysfunction as described in section 5.2. There is a definite murmur and the aortic sound has an unusually low frequency structure. As would be hoped, the replacement of the diagnosed dysfunctional aortic valve has resulted in the disappearance of these pathological signs and the after surgery recording (Figure 5.14b) shows many typical normal valve characteristics. An interesting change in Figure 5.14 that has occurred from abnormal to normal recordings is the transfer of maximum energy component from the first sound in the dysfunctioning case to the second sound in the normal valve case. This energy shift may have resulted due to the new mechanical valve generating more efficient pressure build up prior to aortic valve opening which in turn results in a more explosive and higher intensity aortic sound.

The most significant feature of Figure 5.14 is the fact that replacement of the dysfunctional aortic valve has had a definite effect on both the first and second heart sound. Considering the model in Figure 5.13 in relation to the before and after surgery results in Figures 5.14a and 5.14b it can be seen that the only difference in the sound generation system, between recordings, is the structure and condition of the aortic valve. Modulation of both the first and second sounds, following valve replacement, suggests that it is valid to assume that the structure and condition of the aortic valve effects the whole cardiac cycle of sounds.
Figure 5.14: Time-Frequency distributions of a subject before and after surgery: (a) abnormal native aortic valve; (b) the same subject after replacement of dysfunctioning aortic valve with a metallic prosthesis.
Figure 5.15: Time-Frequency distributions of a subject before and after surgery:
(a) abnormal native mitral valve; (b) the same subject after replacement of
dysfunctioning mitral valve with a metallic prosthesis.
Figure 5.15 shows the time-frequency distributions of the sounds recorded from a subject before and after the replacement of a dysfunctioning native mitral valve with a metallic prosthesis. Due to the very low relative intensity of the second sound in both Figures 5.15a and 5.15b it is impossible to comment on the effect that mitral valve replacement has had on the structure of the second sound. A number of observations can be made about the effect of mitral valve replacement on the first heart sound. As was demonstrated in section 5.4, the heart sounds recorded from a subject with an implanted mechanical valve have definite high frequency components. Hence, it is unsurprising that mitral valve replacement results in an increase in the first sound maximum frequency. In addition to the increase in maximum frequency it can be seen that, after valve replacement, the first sound has become increasingly concentrated about a point in time and frequency. It could be said that the first heart sound, in a state of mitral valve dysfunction (Figure 5.15a), shows characteristics of dispersion with no one dominant component. These changes in the structure of the first sound under constant system conditions (Figure 5.13) supports the hypothesis that mitral valve construction and integrity has a definite effect on the appearance of the time-frequency structure of the first heart sound.

Table 5.5 shows the distribution of the nine descriptive features extracted from the before and after surgery population.

<table>
<thead>
<tr>
<th>Group</th>
<th>N12</th>
<th>NS</th>
<th>f1</th>
<th>f2</th>
<th>f1max</th>
<th>f2max</th>
<th>S1-S2</th>
<th>R-S1</th>
<th>T-S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>aortic before</td>
<td>1/1, 7/2</td>
<td>6/1, 2/2</td>
<td>mean</td>
<td>38.0</td>
<td>46.6</td>
<td>103.9</td>
<td>113.3</td>
<td>352.1</td>
<td>56.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD</td>
<td>7.7</td>
<td>22.3</td>
<td>15.4</td>
<td>36.5</td>
<td>24.4</td>
<td>21.0</td>
</tr>
<tr>
<td>aortic after</td>
<td>8/2</td>
<td>8/0</td>
<td>mean</td>
<td>23.8</td>
<td>38.1</td>
<td>71.0</td>
<td>125.4</td>
<td>236.0</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD</td>
<td>3.4</td>
<td>18.0</td>
<td>30.6</td>
<td>46.1</td>
<td>39.8</td>
<td>27.3</td>
</tr>
<tr>
<td>mitral before</td>
<td>1/1, 7/2</td>
<td>5/0, 3/1</td>
<td>mean</td>
<td>37.1</td>
<td>30.4</td>
<td>119.0</td>
<td>74.3</td>
<td>320.0</td>
<td>72.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD</td>
<td>14.5</td>
<td>5.8</td>
<td>18.9</td>
<td>16.7</td>
<td>63.3</td>
<td>19.0</td>
</tr>
<tr>
<td>mitral after</td>
<td>1/1, 7/2</td>
<td>6/0, 2/1</td>
<td>mean</td>
<td>31.1</td>
<td>27.4</td>
<td>120.0</td>
<td>58.4</td>
<td>268.5</td>
<td>69.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD</td>
<td>11.0</td>
<td>12.4</td>
<td>27.4</td>
<td>12.5</td>
<td>40.5</td>
<td>27.1</td>
</tr>
</tbody>
</table>

**Table 5.5:** Distribution of N12 and NS results together with mean and SD values for the seven T-F parameters for all four before and after surgery population groups. Results for N12 and NS are presented in the form 10/2, indicating that 10 of the group members have 2 sounds.
All subjects with aortic dysfunctioning valves have, before valve replacement, an extra sound in addition to the first and second sound. After valve replacement all these subjects have lost their extra sound. Again, these results point to normality being characterised by a sound cycle having first and second sound components only. This clear distinction between before and after recordings is not shown in the mitral valve case, in particular it can be seen that many of the abnormal mitral subjects, before surgery, do not exhibit extra pathological sounds.

A surprising characteristic of the before and after surgery subjects is that, over the low frequency range considered, there are only a few significant differences between before and after recordings. In all subjects in this population a dysfunctioning native valve was replaced with a metallic prosthesis and hence, as was demonstrated in Table 5.4, a significant change in \( f_{2\text{max}} \) and perhaps \( f_{1\text{max}} \) would have been expected. In contrast to the lack of change in frequency characteristics it can be seen that a number of the time features, in both mitral and aortic cases, show considerable change. This modification to the time features may well be caused by the heart not having adapted to the improved valve performance. In the dysfunctioning valve case the heart can often change its physical characteristics to compensate for inefficient valve operation. With a sudden change in valve integrity, the heart can not return immediately to its normal state and hence, the timing of the valve transitions becomes abnormal.

**5.6 Summary**

In the previous four sections a substantial number of typical examples of the time-frequency distribution of heart sounds have been presented. One particular objective, described in sections 5.2 and 5.3, has been to investigate the potential that time-frequency provides for the identification of abnormal valve subjects. It has been shown, via typical results and the population distributions for a set of descriptive features, that there are definite differences in the time-frequency nature of the heart sounds recorded from normal and abnormal subjects. The results in section 5.2 suggest that the differences between normal and abnormal valves vary depending upon the population being analysed, i.e. the factors that characterise a dysfunctioning aortic valve are different to those that can be used to identify a dysfunctioning mitral valve. This observation supports the conclusion that not only will it be possible to identify valve condition but it will also be possible, using time-frequency sound analysis, to identify the valvular origin of dysfunction.
A particularly interesting and useful result that has been demonstrated by the majority of examples in this chapter is the high degree of consistency of normal valve time-frequency distributions. Irrespective of valve type or construction a subject with two normally functioning valves has heart sounds that exclusively contain significant first and second sounds. The results presented in this chapter support the auscultatory assumption that the appearance of a sound other than the first and second sound is very often an indication of valve dysfunction.

An exception to this consistent appearance of normal valve sounds is shown in the results presented for the bioprosthetic valve population. It was shown in section 5.3 that there are a number of "normal" subjects that show features that would usually be associated with the sounds recorded from a subject with a known dysfunctioning valve. A probable explanation for these inconsistencies in the normal bioprosthetic population is that these subjects are showing signs of dysfunction that are yet to be discovered during routine post-operative monitoring. Such results suggest that the time-frequency analysis of recorded heart sounds offers potential as a tool for the post-operative monitoring of patients with implanted bioprosthetic valves. It should be noted that all subjects included in the normal bioprosthetic population were interviewed prior to sound recording and all indicated that they were in a state of good health. The results presented here suggest that time-frequency techniques can be used as an early indicator of bioprosthetic valve dysfunction and hence provide the opportunity for prompt treatment resulting in reduced patient mortality.

In addition to the objective of identifying valve dysfunction it has been noted that time-frequency analysis presents potential for describing the origins and mechanisms governing the production of heart sounds. Due to the valvular nature of this study, the information available is based around the support or disassociation of sound origins with valve vibration. In section 5.3 it was demonstrated that subjects with normally functioning prosthetic valves have an initial sharp high frequency characteristic to the first sound in mitral cases and to the second sound in aortic cases. These results suggest that the valve integrity has a definite effect on the time-frequency structure of the first and second heart sounds. This valvular influence on the early component of the second sound is supported by the appearance of high frequency components at a similar point in sounds recorded from subjects with implanted aortic metallic prosthesis. Analysis of these metallic valve recordings suggest that valve vibration plays only a very small role in the generation of heart sounds. In subjects with implanted metallic prosthesis the majority of sound energy is concentrated around the 30-40 Hz frequency range. There is no doubt that these major components are generated by the
vibration of system components such as the blood filled ventricles or the great vessels of the heart.

The analysis of sounds recorded from subjects before and after surgery also provide interesting observations on the association of valve closure with specific sounds. In the past, many researchers have made the assumption that the first and second sounds are principally associated with the mitral and aortic valves. The examples shown in section 5.5 quite clearly demonstrate that replacement of a dysfunctioning native valve with a normal metallic valve has an effect not only on the structure of the associated sound but on the whole cardiac cycle of sounds. This result brings into question the validity of the assumption made in past work [47][70] that the first and second sounds are principally associated with the mitral and aortic valves respectively.

The discussions presented above illustrate that time-frequency is a useful tool for the complete description of heart sounds over the whole cardiac cycle. This is in contrast to previous work using frequency only spectral estimation techniques where an assumption of signal stationarity has meant that analysis must be concentrated on a short snapshot of the cardiac cycle. It can be concluded that for complete description of the recorded heart sounds, non-stationary analysis techniques provide a number of distinct advantages over classical spectral estimation techniques.

In comparison to auscultation techniques it can be seen that time-frequency provides a mechanism by which a quantitative analysis of sounds can be made. Traditional auscultatory methods provide a description of the cardiac sound cycle in subjective terms such as high-pitched, low-pitched, blowing and click. A time-frequency analysis tool provides definite quantities that can be used as a measure of valve operation and condition.

From the results in this chapter it is possible to suggest a theory for the origin and generating mechanisms of surface recorded heart sounds:

The time and frequency structure of the cardiac cycle of heart sounds is dependent upon the vibration of the whole cardiac structure. Valve integrity effects the operation of this system and, hence, has an effect over the whole sound cycle. Valve structure and integrity directly effects the initial components of the first and second sounds. Initial first and second sound vibrations contain components that
are a direct result of valve structural vibrations. Efficient closure of an effective valve results in an instantaneous and high frequency initiation to vibration. The majority of first and second sound vibrational energy is generated at low frequency and is independent of valve structure. These components are most probably due to the vibration of myocardial tissue, contained blood and the great vessels of the heart.
Chapter 6

Feature Extraction and Classification

In the previous chapter specific examples of all the data populations collected in the survey were shown. Considering these typical results a number of comments were made about the nature and origin of heart sounds. A major objective of this work is the investigation of the validity of time-frequency signal analysis techniques as a tool for the description and diagnosis of recorded heart sounds. In chapter 5 the conclusion was made that time-frequency techniques can be used to provide a comprehensive description of recorded heart sound data. In addition to this it was also postulated that these detailed descriptions provide diagnostic potential. To validate this claim and to compare results to past work using one-dimensional spectral estimation techniques [58][66], it is essential that an attempt to classify the sounds is made. This chapter describes two methods for the extraction of time-frequency features and the classification of subjects using these extracted features. The data used in this section is limited to the two populations: native valves and aortic C-E bioprosthetic valves. Other data populations were not used in this classification work due to either insufficient subject numbers or the absence of abnormal patient data.

6.1 Feature Extraction

Classically there are two alternative approaches to feature extraction, the extraction of a limited number of descriptive morphological features and the analysis of a large set of possibly less descriptive features to find an optimum feature subset. These two different approaches have distinct advantages. The first method results in the description of the data via features that quite often have physical meaning. For example, in this application we can relate morphological features to classical auscultation extracted features. The analysis of a large set of
features for the identification of an optimal feature set has the advantage that identification and extraction of features is a simple task and classification results are optimised. These two alternative approaches to feature extraction impose different requirements on the time-frequency signal description. The efficient extraction of morphological features requires a clear visual description of the signal and, as was concluded in chapter 4, the CWD is the most suitable transform for this purpose. To determine and extract an optimised feature set, alternative properties are required from the time-frequency signal description. These properties are efficient and complete description of the signal. As was discussed at the end of section 4.7, these alternative requirements may result in the CWD no longer being the most suitable tool for signal description. A description of the extraction of morphological features from time-frequency data is described in section 6.2. Section 6.3 describes the analysis of the suitability of the CWD to optimal feature extraction as compared to the DWT technique described in section 4.2.1. In sections 6.2 and 6.3 results are presented for the classification of native valve condition, native valve position of dysfunction and bioprosthetic valve condition.

6.2 Morphological Feature Classification

The classification stage of an analysis performs the job of assigning a distribution to the known data and assigning a class to the unknown test data. This section concentrates on the extraction of morphological features and the subsequent classification of the heart sound data using these features. The extraction of morphological features is desirable for two main reasons:

1. An objective of this work was to prove that the time-frequency description of heart sounds can be used as a visual aid for the identification of subjects with dysfunctioning cardiac valves. It should be noted that this is in contrast to the aim of the majority of past researchers, of achieving the classification of normal/abnormal subjects in a black-box manner. It is envisaged that the true application of this work would involve the human evaluation of the time-frequency information for a given subject and the classification of subject condition by a human expert. In such an application scenario the human expert would undoubtedly extract morphological features and hence, it is desirable at this initial stage to demonstrate the classification potential of morphological features.

2. At present, auscultation is used successfully in the diagnosis of cardiovascular condition. It has already been noted that auscultation is inherently a time-frequency technique. Hence, using morphological features to perform a classification, it should be
possible to mimic the performance and hopefully the success of auscultation as a tool for heart sound diagnosis.

The next section describes the format of features that are extracted by an experienced auscultator in an attempt to describe sounds in the cardiac cycle. These auscultation concepts are used to guide the definition of a set of morphological features.

6.2.1 Classical diagnostic parameters and morphological features

Many times in this thesis it has been noted that time-frequency techniques have a clear relation to the human auditory system and, hence, auscultation. The advantage of this mimicry is that it is possible to incorporate auscultation knowledge into the procedures used to analyse heart sound time-frequency distributions. The area in which most advantage might be gained is that of feature extraction. By noting and understanding the features extracted using auscultation, we can create a definition for the time-frequency features that will most probably provide useful diagnostic information.

Time-domain features

During auscultation timing information is extracted from the heart sounds in a number of forms. Firstly, direct time information is extracted such as the duration of systole and the heart rate (i.e. the time from the first sound to the second sound and the time between successive first or second sounds). Time information is also used indirectly for tasks such as the labelling of specific sound components. For example, a sound appearing early in diastole after the second sound could be a third heart sound or possibly an early diastolic murmur. Time information is also used to describe and identify pathological murmurs. By comparing the timing of a murmur with respect to the two major sounds the murmur is characterised as either systolic or diastolic. This time description of a murmur is further refined by the addition of the prefix pan, early, late or mid to the diastolic or systolic label. The categorisation of a sound as a murmur is influenced by its frequency characteristics but is also dependent on its relatively long duration in time compared to other cardiac sounds.
Chapter 6 : Feature Extraction and Classification

Frequency-domain features

The term frequency is rarely used in the literature as a descriptor of sounds, most often sounds are described in terms of their pitch. Commonly, the first and second sounds are described as high-pitched, compared to the low-pitched third and fourth sounds. Other sounds that are commonly given a high-pitched description are pathological snaps and clicks. Murmurs are often given a frequency description they can be either low, medium or high-pitched. In general murmurs have a higher pitch than other cardiac sounds and hence, can be easily identified.

From the above descriptions it is clear that both time and frequency information is used during auscultation to identify and describe heart sounds. It is clear from the medical literature [28][119][29] that the description of these sounds does vary and as such is highly subjective. In fact, it is quite clear that analysis of sounds is very much an art that is learnt by practitioners via experience.

By studying the literature it can be seen that in the majority of cases analysis and diagnosis of sounds is based on time information. Frequency information is very rarely a major factor in the identification of a pathological condition. This concentration on the time information and disregard for frequency information is quite probably due to the auditory system's very poor frequency resolution. The human expert using auscultation is unable to resolve discrete frequency sound components in the limited band over which heart sounds are concentrated.

Using the above information and past experience with auscultation it is possible to design a number of simple features that will offer diagnostic potential. The first step of an auscultation diagnosis is the identification of which significant sounds are present in the cardiac cycle. It is well-documented that in normal adults only the first and second sounds are present. Considering this, it was felt that for diagnosis to be successful, information concerning the appearance of the first, second and other pathological sounds should be collected. Although frequency information is not commonly used in auscultation techniques, it was felt that for the complete definition of the first and second sound characteristics a number of frequency features should be included. As already mentioned above, time-information is used extensively in the auscultation process. Concurrent with this, three simple time parameters were extracted from the sound cycle using the ECG signal as a time reference. The nine
Chapter 6: Feature Extraction and Classification

extracted features are given below in Table 6.1. These features were extracted using the techniques described in section 5.2. The binary features N12(b) and NS(b) were determined from the states of N12 and NS in the original features of Table 5.1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N12(b)</td>
<td>Are S1 and S2 both present? (1/0)</td>
</tr>
<tr>
<td>2</td>
<td>NS(b)</td>
<td>Are sounds other than S1 and S2 present? (1/0)</td>
</tr>
<tr>
<td>3</td>
<td>f1</td>
<td>Frequency of S1 maximum intensity (Hz)</td>
</tr>
<tr>
<td>4</td>
<td>f2</td>
<td>Frequency of S2 maximum intensity (Hz)</td>
</tr>
<tr>
<td>5</td>
<td>f1max</td>
<td>Highest S1 frequency (Hz)</td>
</tr>
<tr>
<td>6</td>
<td>f2max</td>
<td>Highest S2 frequency (Hz)</td>
</tr>
<tr>
<td>7</td>
<td>S1-S2</td>
<td>Time between features f1 and f2 (ms)</td>
</tr>
<tr>
<td>8</td>
<td>R-S1</td>
<td>Time from ECG R wave to feature f1 (ms)</td>
</tr>
<tr>
<td>9</td>
<td>T-S2</td>
<td>Time from ECG T wave to feature f2 (ms)</td>
</tr>
</tbody>
</table>

Table 6.1: Nine Morphological time and frequency features.

6.2.2 Overall classification procedure

The simplest form of classification problem involves the assignment of the input data to a two class distribution. This two class normal / abnormal classification has been the aim of all past work attempting to analyse heart sounds. In this work an attempt has been made to perform a more complex diagnostic analysis by identifying the origin of valve dysfunction. Achieving this classification is further complicated by the varying format of the features in Table 6.1. Features 1 and 2 have a binary result whereas features 3-9 all take on continuous values. Taking into consideration these complications it is important to define an overall structure to the classification procedure. Classification is performed in two stages. First, the data is assigned to either a normal or abnormal class and secondly, the identified abnormal data is separated into two further aortic, mitral sub-classes. At each of these two stages classification is achieved by combining the results of the classification using the binary features 1, 2 and the results of classification using the continuous features 3-9. Figure 6.1 shows a graphical representation of the format of this overall classification structure.
The illustration in Figure 6.1 has two major decision blocks, the binary feature classifier and the continuous feature classifier. The first stage of classification generates a decision as to the condition of the subject’s valves. At this stage the results of the two independent binary and continuous classifications are combined by an OR function (i.e. if either the binary decision indicates dysfunction or the continuous decision indicates dysfunction then a final decision of valve dysfunction is made). The identified abnormal data is again analysed in the second stage of Figure 6.1 in an attempt to identify if dysfunction originates at the aortic or mitral valve. As is demonstrated later in section 6.2.5 the binary features show a very poor performance in identifying the origin of dysfunction and hence, the binary decision is omitted from the second stage of the overall decision process.

This chapter includes the analysis of two populations of data, native valve subjects and the aortic C-E bioprosthetic valve subjects. As can be seen from the distribution of these populations in Figure 3.8 only the native valve data can be used together with the full structure of Figure 6.1. In the bioprosthetic valve case, all the subjects are known to have aortic valves and hence, only the first stage of normal / abnormal classification can be performed. The following two sections describe the techniques used to implement the two main decision blocks in Figure 6.1, the binary and continuous feature classifier respectively.
6.2.3 Binary feature classification

Features 1 and 2 in Table 6.1 both result in a binary decision, i.e. they are either yes or no (1/0). The first stage of classification using these binary features is to estimate the probability of each feature combination for each possible class. By applying unknown data to these probability distributions it is possible to estimate class membership based upon the highest probability class assignment. A single feature situation with a binary probability distribution for a two class problem would take the form shown in Figure 6.2a. A two feature situation again for a two class problem would take the form shown in 6.2b.

![Figure 6.2: Binary feature probability distributions: a) single feature case; b) two feature case.](image)

If we consider the distributions in Figure 6.2 to have been generated under a situation of supervised learning we can estimate the class of unknown data by comparison with these distributions. In the simple single feature case an unseen test subject may have a feature value of 1. It can be seen from Figure 6.2a that the probability of this test data belonging to class 1 is higher than the probability of its membership being in class 2, hence, we allocate the test data to class 1. This decision procedure can be extended to any dimension and can be expressed formally as:

\[
\text{class} = n \quad \text{if} \quad p(x|c_n) > p(x|c_i) \quad 0 \leq i \leq N \quad i \neq n \quad (6.1)
\]
Where $x$ is the test subject, $p(x|c)$ is the conditional probability of the occurrence of $x$ given $c$ and $N$ is the number of data classes. The expression in (6.1) is the general form of the Bayes' rule [120].

### 6.2.4 Continuous feature classification

This section includes the descriptions of two classification procedures. The first is the statistical Bayes classification scheme and the second is the $K$-nearest neighbour deterministic classification scheme.

#### Statistical classification scheme

To perform a statistical classification using continuous features it is necessary to estimate a probability density function for each possible subject class. As in the binary feature classification scheme, a knowledge of these probability density functions for the known data provides us with a guide to the true class of any unknown data.

In the binary data cases, shown in Figure 6.2, it can be seen that although the levels of probability are not known, the general step structure of the distributions will always occur. In the continuous case, there is no such a-priori knowledge of the underlying distribution shape. Hence, the distribution needs to be estimated. An idea of distribution shape can be found by generating a histogram of the data and estimating its shape. Generating a histogram of the data that provides a clear indication of the data distribution requires an extensive amount of data. It can be seen from Figure 3.8 that in the native valve population there are only 20 subjects per class and in the C-E bioprosthesis valve population there are only 11 subjects per class. These relatively small class populations mean that estimation of the shape of probability distributions can not be effectively performed by generating a data histogram. As an alternative to presenting the data as a histogram, it is possible to generate an idea of data probability distributions by representing the data as a scatter plot and inspecting the concentration of points. Figure 6.3 shows a scatter plot for all seven features for normal and abnormal classes in both the native valve population and the C-E bio-prosthetic valve population.
Figure 6.3 continued over page
Figure 6.3: Distribution of the seven continuous features for both native and bioprosthetic valve subjects. (a)-(g) represent features 3-9 in Table 6.1. The distribution for normal native valves is printed at NN, abnormal native valves at AN, normal bioprosthetic at NB and abnormal bioprosthetic at AB.

In the plots of Figure 6.3 there seems to be no one clearly dominant format to the distributed feature points. The distribution of normal native subjects in Figure 6.3a shows a concentration about a feature value of 40 with the outlying points becoming more sparse as the feature value spreads. This arrangement of points suggests these features may conform to a normal distribution. The distribution of abnormal native subjects in Figure 6.3d again exhibits a single concentration of points but outlying points only occur as the feature value increases. With limited amounts of data it is impossible to accurately estimate data distributions but in this case, to allow a statistical classification of the data to be performed, an assumption of normal distribution is made. The validity of an assumption of distribution shape may be tested using a technique referred to as the goodness of fit \( \chi^2 \) test [121]. The test is performed by measuring the difference between the true distribution and the assumed normal distribution. As already noted, due to small population sizes used in the study, it is not possible to accurately estimate the true distribution of the data set, hence, the \( \chi^2 \) test can not be applied.

Making the assumption of data being normally distributed an attempt to measure the data distributions and perform a classification can be made using the Bayes decision rule. The Bayes decision rule has already been used in section 6.2.3 for the classification of binary features. In section 6.2.3 a decision was made in order perform a class decision given a set of measurements and the probabilities that these measurements belong to a known class. The highest probability of class membership provides an indication of the likely class of the test.
data.

As in the binary feature case (Figure 6.2), to perform a classification of test data the shape of the probability distributions for each data class must be estimated from a set of training data. To specify the shape of a multivariate normal distribution of the \( n \) features in an \( n \)-dimensional space, the covariance matrix (\( \Sigma \)) must be calculated:

\[
\Sigma_{ij} = \frac{1}{N} \sum_{k=1}^{N} (f_i(k) - \bar{f}_i)(f_j(k) - \bar{f}_j)
\]

\[i = 1,2,..n \quad j = 1,2,..n\]  

(6.2)

where \( N \) is the number of subjects in the training set, \( n \) is the number of features and \( \bar{f}_i \) is the mean value for the \( f_i \) feature. For each possible population class a covariance matrix is generated and probability of class membership for a test subject \( x \) is calculated as:

\[
p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu)}
\]

(6.3)

where \(|\Sigma|\) is the determinant of the covariance matrix \( \Sigma \), \( x \) is the unknown feature vector, \( \mu \) is the mean feature vector of all the known subjects and \( \Sigma^{-1} \) is the inverse matrix of \( \Sigma \).

The most important factor in (6.3) is the covariance matrix which defines the distribution of the data features and their interrelationship. It is the structure of \( \Sigma \) that governs the separability of the subject set and hence, the success of classification. Calculating the covariance matrix for the class of normal native valves results in the 7 by 7 matrix shown in Figure 6.4.

A value for the determinant of the covariance matrix \(|\Sigma|\) can be calculated as:

\[
|\Sigma| = \lambda_1 \times \lambda_2 \times \lambda_3 \times \ldots \times \lambda_n
\]

(6.4)

where \( \lambda_i \) represents an eigenvalue of the matrix. The determinant of the covariance matrix for the normal native valve population is \( 6.599976 \times 10^{-13} \). It can be seen from (6.3) that such a small value for \(|\Sigma|\) will cause \( p(x) \) to take on extremely large values. These large values result in rounding errors that in turn result in impractical estimation inaccuracies. It can be seen from (6.4) that a very small matrix determinant will occur if a number of the matrix eigenvalues are also small. The eigenvalues for normal native valve covariance matrix are shown in the graph of Figure 6.5 in order of descending magnitude.
Figure 6.4: Covariance matrix for the seven continuous features extracted from the time-frequency distributions of normal native subjects.

<table>
<thead>
<tr>
<th></th>
<th>0.0368</th>
<th>0.0161</th>
<th>0.0197</th>
<th>0.0012</th>
<th>-0.0004</th>
<th>0.0005</th>
<th>0.0068</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0016</td>
<td>0.0296</td>
<td>0.0039</td>
<td>0.0196</td>
<td>-0.0046</td>
<td>0.0199</td>
<td>0.0032</td>
<td></td>
</tr>
<tr>
<td>0.0197</td>
<td>0.0039</td>
<td>0.0273</td>
<td>0.0020</td>
<td>-0.0003</td>
<td>-0.0044</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td>0.0012</td>
<td>0.0196</td>
<td>0.0020</td>
<td>0.0283</td>
<td>0.0005</td>
<td>-0.0031</td>
<td>-0.0007</td>
<td></td>
</tr>
<tr>
<td>-0.0004</td>
<td>-0.0046</td>
<td>-0.0003</td>
<td>0.0005</td>
<td>0.0077</td>
<td>-0.0263</td>
<td>0.0094</td>
<td></td>
</tr>
<tr>
<td>0.0005</td>
<td>0.0199</td>
<td>-0.0044</td>
<td>-0.0031</td>
<td>-0.0263</td>
<td>0.1354</td>
<td>-0.0143</td>
<td></td>
</tr>
<tr>
<td>0.0068</td>
<td>0.0032</td>
<td>0.0026</td>
<td>-0.0007</td>
<td>0.0094</td>
<td>-0.0143</td>
<td>0.0346</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.5: Eigenvalues of the covariance matrix calculated for the features extracted from normal native valve subjects.
It can be seen that eigenvalue 0 is two to three times larger than any other value and over one hundred times greater than the smallest eigenvalue. This contrast between eigenvalues results in what is commonly referred to as an ill-conditioned covariance matrix. The ill-conditioning of the covariance matrix suggests that it is rank deficient. This rank deficiency, in turn, suggests a redundancy in the matrix and a redundancy in the feature set. There are two alternative solutions to this problem of covariance matrix rank deficiency: reduce the number of features or find better features.

From inspection of Figure 6.3 it can be seen that no single feature provides a clear indicator of valve condition. Reducing the number of features used in the classification process, although resulting in a better conditioned matrix, will most probably reduce the success of subject classification. The alternative to reducing the number of features used is to extract a better set of discriminative features. The guided extraction of features from the time-frequency plots is possible using techniques such as principle component analysis [122]. This section of the thesis describes the investigation of the suitability of time-frequency distributions as a visual aid to the diagnosis of cardiac valve condition. To this end it is an important factor of the classification that the extracted features are morphological. It is envisaged that other features (points on the time-frequency plane) will also provide discriminative information, but these features will not relate to the analysis that an experienced auscultator would perform. The extraction of an optimised set of time-frequency features is investigated later in section 6.3. Hence, it is felt that an attempt should be made to use the simple descriptive features shown in Table 6.1 to classify the native and C-E bioprosthetic populations.

As can be seen in (6.2) and (6.3) the covariance matrix is calculated in advance of utilising the assumption of normal distribution to estimate class membership. On considering this relationship it is clear that the assumption of a normal distribution of features has not contributed to the rank deficiency of the covariance matrix. However, it should be noted that if rank deficiency of the covariance matrix had not occurred classification may still have been problematic due to the questionable validity of the assumption of normal feature distribution.

Although these results suggest that a classification using the features of Table 6.1 will be problematic, it should be noted that the lack of success of this statistical technique, and quite possibly other statistical techniques, is based upon the assessment of the information content of the extracted features. Although the information content of features is significant, it is not the ultimate aim of classification to describe the data, rather it is to perform a discrimination between subject groups. Due to the concentration of statistical techniques upon the
information content of classification features and the inherent correlation often encountered when using morphological features, it becomes clear that to perform classification an alternative decision mechanism is required. Such an alternative is provided by deterministic classification techniques. The next section describes the theory, application and results of using the K-nearest neighbour deterministic classification technique [122].
Deterministic classification scheme

A highly logical approach to the problem of classification is to assign unknown data to the class in which its nearest neighbours reside. Consider the situation in Figure 6.6.

Allocation of the subject A to a class is a trivial problem. A's nearest neighbour is obviously a member of class +. In fact, for the situation shown in Figure 6.6, any practical classification scheme could not fail to find the correct class for A. A less trivial problem occurs when the training population contains what may be referred to as rogue subjects (Figure 6.7). In this case it is possible for A's nearest neighbour to be in either of the two classes. To overcome this rogue value problem the nearest neighbour measurement is extended to a number of nearest neighbours. In the example of Figure 6.7, if classification is performed by calculating A's five nearest neighbours and allocating class membership to the majority represented class, the two rogue values will never cause a misclassification. Hence, this K-nearest neighbour classification technique provides an efficient and simple classification scheme that avoids the statistical problems that small data sets and morphological features can present.
There are two practical factors to consider when using the K-nearest neighbour scheme: the measure used to calculate the distance between subjects and the number of neighbours that should be used to avoid rogue subject problems. Measuring the distance between subjects can be performed in a number of ways [120] and common distance measures include the Hamming distance, Euclidean distance, city block distance and square distance. Of these the Euclidean distance is the most often used and provides a true distance between feature vectors. The other distance measures produce approximations to the true subject separation with the advantage of reduced calculation. In this application we are dealing with limited data and a relatively small feature set, hence, the exact Euclidean distance measure has been used. The Euclidean distance between two feature vectors $x$ and $y$ can be calculated as:

$$d(x, y)_{\text{euc}} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$  \hspace{1cm} (6.5)

Where $n$ is the number of individual features in the vectors $x$ and $y$ (the value of $n$ also determines the dimensionality of the feature space).

Another important factor to consider when using the K-nearest neighbour classification scheme is the number of neighbours that should be used to decide upon the class of test data. In a two class problem it is desirable to measure an odd number of neighbours in order to ensure a decision is made. Considering the number of neighbours that should be included in the decision it has been shown previously [123] that increasing the value of $K$ does not
produce a simple increase in classifier performance. In fact, a general rule for the value of K is that it should be equal to the square root of the number of patients in the training set [124].

6.2.5 Classification of native valve subjects using morphological features

In the previous two sections, descriptions for both binary and continuous classification procedures have been provided. In both instances it has been shown that the identification of the class of an individual test subject is performed by comparison with a group of known training data. The training and testing of the binary and K-nearest neighbour schemes would be best performed using two separate sets of training and test data, ideally of comparable size. This ideal setup requires a data set that is of substantial size (100+). In this case and in all past surveys it is clear that due to logistic reasons a sample set of this size is not available. With the small sample sizes available it is important that the available data is used efficiently. Data could be maximised by using all data as both training and test data, but this technique clearly produces over estimated results due to what can be described as fitting the result to the available data not the population it represents. A compromise between these two extremes is achieved in this work by using the leave-one-out method, the classifier is trained on all but one of the available data samples which is employed as the test data. This training and test procedure is repeated with each data sample as test data in turn until all data has been tested. Using this method the number of available training subjects is maximised at \( N - 1 \) and the test data retains the desirable property of being new unseen data.

Applying the binary and continuous classification techniques described above to the native valve population in an attempt to classify valve condition produced the results shown in Table 6.2. In accordance with the discussion in section 6.2.4 classification using the continuous features was performed using a seven nearest neighbour analysis.
Table 6.2: Continuous and binary classification results for the identification of normal and abnormal native valve subjects.

<table>
<thead>
<tr>
<th>subject N°</th>
<th>81</th>
<th>80</th>
<th>79</th>
<th>78</th>
<th>77</th>
<th>76</th>
<th>75</th>
<th>74</th>
<th>73</th>
<th>72</th>
<th>71</th>
<th>70</th>
<th>69</th>
<th>54</th>
<th>53</th>
<th>52</th>
<th>51</th>
<th>50</th>
<th>49</th>
<th>48</th>
<th>47</th>
<th>46</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>known class</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>binary classification</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>K-neighbour classification</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| subject N° | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 | 33 | 22 | 21 | 20 | 19 | 18 | 16 | 15 | 14 | 13 | 12 |
|------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| known class | 2  | 2  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |
| binary classification | 2  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |
| K-neighbour classification | 1  | 2  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 2  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |

In previous work attempting to analyse heart sounds there has been a correlation of the origin of valve dysfunction with the areas in which recordings have been made. Section 3.3 of the thesis has a full description of the commonly used mitral and aortic recording sites. In the data set used in this section we are attempting to classify valve condition irrespective of the valve position. Hence, in contrast to previous work, a common mitral site recording will be used for all three dysfunctioning aortic, dysfunctioning mitral and normal valve subject populations. This consistency of recording site is essential to prevent any bias due to the affect the chest thorax system may have on sounds heard at various points on the chest surface.

The subject numbers used in Table 6.2 were arbitrarily allocated as a means of anonymous patient identification. The subjects in Table 6.2 belong to one of two known groups, either normal native or abnormal native valves. In Table 6.2 these two groups have been labelled class 1 and 2 respectively. From these results it can be seen that the binary classification scheme has correctly identified all 20 of the normal subjects and 16 of the 25 abnormal subjects. The K-nearest neighbour scheme has correctly identified 19 of the 20 normal subjects and 19 of the 25 abnormal subjects. The binary and continuous results can be combined such that a subject is classified as abnormal if either individual result indicates abnormality. Such
a combination results in 43 of the 45 subjects being correctly classified. By calculating these results in terms of percentages, it can be seen that the individual binary and continuous schemes produce 80% and 84% correct classification respectively and the combination of binary and continuous results produces a final 96% correct classification.

Although it is important to be able to determine the integrity of a patient's cardiac valves, these results are only the first stage of the analysis a patient must undergo. To treat any form of dysfunctioning valve, it is necessary to evaluate more completely the classified abnormal cases. In particular it is desirable to know the origin of dysfunction. The abnormal population shown in Table 6.2 is composed of 13 subjects with known dysfunctioning aortic valves and 12 subjects with known dysfunctioning mitral valves. Using this information, it is possible to perform a more complete diagnosis by attempting to assign a classification of origin to these identified abnormal valves. In a practical application of these techniques only the identified abnormal subjects would be passed forward for further scrutiny. In fact, due to the a-priori knowledge of valve function this second stage is applied to all known abnormal subjects rather than the identified abnormalities in Table 6.2. Results for binary and continuous feature classification of the origin of dysfunctioning native valves are shown in Table 6.3. Classification using the continuous features was performed using a five nearest neighbour analysis.

<table>
<thead>
<tr>
<th>subject N°</th>
<th>81</th>
<th>80</th>
<th>79</th>
<th>78</th>
<th>77</th>
<th>76</th>
<th>75</th>
<th>74</th>
<th>73</th>
<th>72</th>
<th>71</th>
<th>70</th>
<th>69</th>
<th>54</th>
<th>53</th>
<th>52</th>
<th>51</th>
<th>50</th>
<th>49</th>
<th>48</th>
<th>47</th>
<th>46</th>
<th>45</th>
<th>44</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>known class</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>binary classification</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>?</td>
<td>?</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>K-neighbour classification</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: Binary and continuous classification results for the identification of aortic and mitral native dysfunctioning valve subjects.

This second test set contains only dysfunctioning native valve subjects (i.e. subjects 43-54 and subjects 69-81, see Table 6.2). Although these subjects are the same as those in Table 6.2 they have been reassigned to new classes. In Table 6.3 class 1 represents subjects with known dysfunctioning aortic valves and class 2 represents subjects with known dysfunctioning mitral valves. Results of the binary classification scheme can take one of three forms: class 1,
class 2 or class "?". The symbol "?" refers to when the probability of membership is equal for both classes and hence, no decision can be made. From Table 6.3 it can be seen that the continuous scheme has correctly classified 16 of the 25 dysfunctioning subjects but the binary scheme has only correctly identified 9 of these subjects. Summarising these results, the binary and continuous schemes produced 36% and 64% correct classification respectively. If a random allocation of class had been performed, correct classification would be in the region of 50%. Considering this, it can be seen that the continuous scheme has not significantly beaten chance and the binary results are in fact useless. The failure of the binary features to provide any indication of the origin of valve dysfunction renders them redundant and hence, no combination of binary and continuous feature results is performed. A possible reason for the poor performance of the continuous feature scheme is presented by Foley[125]. In this work it is presented that up to a point the performance of a classifier increases as the number of discriminant features used in the classification increases. Past the given point further increases in feature set size results in a deterioration in classifier performance. It is proposed in [125] that if the ratio of class size ($N$) to the number of features ($L$), i.e. $N/L$, is lower than 3 then classification results will not reflect the true classifier performance. In the scheme to identify dysfunctioning native valve origin, $N/L$ is equal to 1.714 which, from the work in [125], suggests unacceptable classification errors. To increase this ratio to 3, at which point true classifier performance should become apparent, 3 of the continuous features must be discarded.

Having identified the need to select a subset of features, the question arises as to which features should be kept. There are traditionally two answers to this question: keep the four features that provide the most information about the data or keep the four features that are most discriminative. The first solution proposes that if signal information is conserved, it will be possible to more clearly describe and hence, differentiate class members. This assumption is not always valid and in certain situations it can be shown that this conservation of signal information is detrimental to the ultimate aim of class separation [122][126]. The alternative to this information conservation technique is to measure the performance of features with respect to their discriminative ability. As the main aim of feature selection is to allow class separation it would seem logical to select a feature set that maximises classification performance. The disadvantage of this selection of features based on their discriminative ability is that in situations where large data and feature sets are being analysed computation requirements can be high.
In this example, the task to be performed is to find the best combination of four features from a total of seven original features. This combination problem is often represented by the notation \( ^7C_4 \) and referred to as the binomial coefficient. In the general case, \( ^nC_r \), an expression for the number of possible combinations of \( r \) features from the \( n \) total features takes the form:

\[
^nC_r = \frac{n!}{(n-r)!r!}
\]

where \( n \) is the total number of features, \( r \) is the feature subset size and \( n! \) represents \( n \) factorial. In this case we have \( n = 7, r = 4 \) and hence 35 possible combinations. With such a small number of possible combinations, computation is not a problem and feature selection can be performed on the ideal basis of the combination of four features that produce the highest classification result.

The continuous classification scheme was implemented 35 times for each possible combination of four features. It was found that 12 of the 35 combinations performed as well as the original 7 features and a combination of features 4, 6, 7 and 9 resulted in a maximum 72% correct classification.

### 6.2.6 Classification of aortic C-E bioprosthetic valve subjects using morphological features

In the C-E bioprosthetic valve population there are 12 normal valve subjects and 11 abnormal valve subjects. Taking into consideration the size of these classes the K-nearest neighbour scheme was performed with respect to 5 neighbours. As in the last section, continuous feature classification was performed for an optimised 4 feature subset. Results for binary and continuous feature classification of the condition of aortic C-E bioprosthetic valves are shown in Table 6.4.
Summarising the results in Table 6.4, the binary and continuous schemes produced 65% and 74% correct classification respectively. Combining binary and continuous results, such that only if both show abnormality then classification is assigned abnormal, gives a final 83% correct classification rate.

6.3 Optimised Feature Classification

The previous section has described the diagnosis of native and bioprosthetic populations via the extraction of morphological features. It was noted that the extraction of morphological features was desirable because of their clear relationship with the mechanisms of sound production and the ability to utilise present auscultation knowledge. It was also shown that morphological feature sets can suffer from dimensionality redundancy and inefficient discriminative performance. An alternative to performing classification using morphological features is to extract a whole range of features that contain most, if not all, of the signal information and find a subset of these features that produce maximum classification. This subset feature extraction and optimisation has been employed in past work [58]. A similar selection of an optimised feature set can be performed for time-frequency sound descriptions. In chapter 4 it was concluded that the CWD is the most suitable analysis tool for the visual time-frequency description of heart sounds and hence, the manual extraction of morphological features. As mentioned in section 4.7 and earlier in this chapter, the determination of an optimum feature set presents a number of different requirements. Due to this change in the requirements of the signal time-frequency description, a reassessment of the suitability of the CWD to this task is presented.

Table 6.4: Binary and continuous feature classifications of aortic C-E bioprosthetic valve subjects.
In the extreme case, it would be valid to consider each pixel of a CWD generated from the analysis of a recorded sound as a collection of individual pixel features. Such an assumption for the data used in this thesis which is (with the exclusion of metallic valve recordings) analysed over 1024 ms with the data sampled at 500 Hz, would result in $512 \times 512 = 262144$ features. In accordance with the argument given in [125] true classification rates will only be obtained for feature subsets of size seven or less. Considering (6.6) with $n = 262144$ and $r = 7$, it can be seen that an enormous number of possible seven feature combinations exist. The size of this problem renders selection of the optimum combination impossible to compute. Reduction of this problem to a practical size can only be performed by reducing the inherent dimensionality of the time-frequency distributions. The excessive dimensionality of the complete time-frequency plot points to another advantage of morphological features; specifically, that large amounts of signal information can be concentrated in only a few features and hence, dimensionality reduction is inherent in the extraction of morphological features.

A very simple method for data dimensionality reduction is presented in [122]. It is suggested that features with high correlation should be combined and hence, dimensionality reduction is achieved with minimal loss of information. The simplest form of feature combination is to replace the features to be combined by their average value. In the time-frequency CWD plots, it is logical to assume that pixels (features) close on the time-frequency plane will, in the majority of cases, exhibit a high mutual correlation. Dimensionality reduction could be achieved by dividing the time-frequency plane into an array of localised blocks which are then averaged to produced a reduced dimension time-frequency representation.

Breaking the time-frequency plane into an array of averaged blocks reduces the effective resolution of the time-frequency representation. It can be seen that the previous requirement of a high resolution description of the time-frequency spectrum of sounds has resulted in considerable dimensionality problems and only by abandoning this high resolution requirement can optimised feature set selection be performed.

An alternative description of the non-stationary spectral properties of a signal can be produced using the DWT. The DWT suffers from poor visual interpretation but does provides a number of definite advantages as an optimised feature extraction technique. Using the DWT tree structure of Figure 4.13 implemented using compactly supported orthonormal wavelets, it is possible to break the signal into a set of non-redundant information preserving packets. This reduction of the signal into an orthonormal basis set provides the advantage that the non-stationary structure of the 512-point heart sound data can be represented by 512 time-
scale coefficients rather than the 262144 time-frequency coefficients of the CWD. Orthonormality of the wavelet bases also provides the advantage that the representation is redundancy free and signal information is preserved completely.

Using the DWT as a feature extraction technique requires an amount of careful pre-processing of the signal to ensure consistency of results. The first factor to consider is the synchronisation of the data in time. The data extracted for each population subject was removed arbitrarily about a typical cycle. This arbitrary signal extraction results in a data set that is not synchronised in time. As the analysis tool to be used is the DWT, a time-scale transform, it is obviously sensitive to these asynchronous signal characteristics. Time synchronisation of the data can be achieved by shifting the data such that the R wave of the ECG of each population subject lines up. This shifting operation results in various start and stop times to the time-domain signal. To again preserve consistency, it is important that the WT is applied to data of a consistent length with common terminating points. This re-alignment is achieved by truncating the signal and padding it with zeros to 512 points. Padding the data with zeros creates the problem of a glitch at the padding edge. This is removed by windowing the available data after time shift and before zero-padding. The window function is defined as a modified Hanning window as in (5.1) and (5.2) such that beat length is 384 points and data length is 512 minus the zero padding length. Figure 6.8 shows an illustration of this signal pre-processing.

It can be seen from Figure 6.8 that this process of time synchronisation results in a portion of signal being discarded. When typical sound cycles were extracted from the complete data recordings the full cycle of sound was extracted such that the first and second sound complexes were centralised. This centralisation of the extracted sounds resulted in a degree of time synchronisation and hence, the amount of signal discarded, in the majority of cases, was a small percentage of the whole cycle. It should also be noted that due to the average beat period for the recordings being approximately 800 ms and the extracted data being 1024 ms in length, the signal ends effected by windowing was primarily redundant data.
Figure 6.8: Illustration of the time-domain signal preprocessing of heart sound data to achieve time synchronisation prior to application of the DWT. $\bar{R}$ represents the average time of the R wave of the ECG for the whole population.

Although the DWT represents the data in 512 efficient orthonormal packets, this is still a very large feature set and finding the optimum four and seven feature subset combinations is still impractical. This means that for classification to be achieved some form of initial reduction of the feature set must be performed. This reduction is achieved by measuring the individual performance of the full 512 features in the classification of the data. Using the results of these individual measures, a significant set of what will be called the "top performers" were chosen for further dimensionality reduction via an exhaustive subset combination search. This scheme works on the assumption that features that are individual bad performers will, even if used in combination with other features, not provide useful discriminative information. This assumption will have its exceptions but these will undoubtedly be a very small percentage of the rejected features. If the number of top performers, from which an optimum set is chosen, is kept high, it is valid to assume that the performance of the chosen optimum sub set will be close, if not equal, to the true optimum performance.
6.3.1 Feature subset selection and classification of native valve patient data

Analysing each of the 512 DWT coefficients for the whole native valve population using a K-nearest neighbour analysis performed for 7 neighbours results in a whole array of classification performances. The graph in Figure 6.9 shows a histogram of the number of features at each level of correct classification for the top 83 performers.

![Histogram of top 83 performances](image)

**Figure 6.9: Distribution of the top 83 classification performances for native valve subjects.**

Inspecting Figure 6.9 and stipulating that a feature is only to be labelled a top performer if more than 30 of the 45 native valve data population are correctly classified (67% correct classification), it can be seen that there are 56 top performing features. It should also be noted that the individual top performer with 37 correct classifications attained, on its own, a respectable 82% correct classification rate.

Although the DWT derived features have been extracted blindly, it is possible to comment upon the relationship of the 56 top performing features with respect to the cardiac sound cycle. This comparison is achieved by illustrating the position of each of these features on the time-scale plane. Figure 6.10 shows the time-scale plane where increasing feature performance is represented by an increasing level of grey scale. Also shown is the average...
position of the R wave of the ECG (R̅) and the average position of the T wave of the ECG (T̅) for the whole population.

![Diagram showing time-scale positioning of the 56 top performing native valve features.]

**Figure 6.10:** Time-scale positioning of the 56 top performing native valve features.

From the inspection of Figure 6.10 a number of points can be made about the position and concentration of the top performing features. Firstly, it can be seen that there is a definite absence of any high-scale coefficients as top performers. Making a tentative connection between scale and frequency, this lack of high-scale performers may indicate that low frequency information is common to both normal and abnormal native valve sounds. Alternatively, it is possible, due to the low time resolution that the WT exhibits at low frequency, that the WT coefficients at these high-scale values do not have the resolution required to differentiate valve condition. Taking into consideration that the R wave of the ECG, in general, occurs at the onset of the first sound and the T wave of the ECG, on average, appears 91 ms prior to the second sound main component it can be seen that the large majority of useful features appear during the period of systole. In particular, there is a concentration of high frequency features almost exclusively between this period. This would seem to be in agreement with the results shown in chapter 5 where typical results suggest that the occurrence of midsystolic high frequency sound components are often associated with native valve dysfunction.
In an attempt to characterise the usefulness of these features it is instrumental, as was done in the case of morphological features, to analyse the dimensionality of the covariance matrix. After generating the covariance matrix of these 56 features as defined in (6.2) the associated eigenvalues can be evaluated. Figure 6.11 shows the 56 covariance matrix eigenvalues arranged in order of magnitude.

![Eigenvalues graph](image)

**Figure 6.11: Eigenvalues of the covariance matrix generated from the 56 top performing DWT features extracted from native valve recordings.**

The most obvious feature in Figure 6.11 is the distinctive knee that appears between eigenvalue 18 and 19. This sudden drop in the size of eigenvalues strongly suggests the 56 top performing features are rank deficient to a true dimensionality of 18. This relatively high value for covariance matrix dimensionality suggests that there are a number of these top performing features that are uncorrelated. This result also suggests that there may well be a subset of these features that will improve upon the 82% individual best feature performance.

- Having identified the 56 individual features that provide the most potential, the next stage of the analysis involves the investigation of the classification performance of subsets of these features. Investigation of an optimum subset of the 56 top performers for the identification
Chapter 6: Feature Extraction and Classification

of native valve condition was performed for 2, 3 and 4 feature set sizes. For each subset size, a whole range of performances were achieved. The results of interest were those combinations that achieved the highest correct classification rates. The maximum number of correct classifications for the total 45 subject set were 41, 43 and 45 for the subset group sizes 2, 3 and 4 respectively. A summary of these results in terms of percentage correct classifications is presented in Table 6.5.

<table>
<thead>
<tr>
<th>subset size</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>maximum classification success</td>
<td>41 (91%)</td>
<td>43 (96%)</td>
<td>45 (100%)</td>
</tr>
</tbody>
</table>

Table 6.5: Optimised subset performance for classification of native valve condition.

As can be seen, using 4 sub-features of the 56 top performers it is possible to correctly classify the condition of all 45 subjects of the native valve population. From the argument in [125] it can be seen that it would have been valid to analyse the performance of 5, 6 and 7 feature subsets and it is quite possible that this complete analysis would have resulted in a large number of further subset combinations that completely classify the native valve population.

As in the morphological feature classification described in section 6.2 the abnormal native valve subjects can be sub-divided into aortic and mitral classes. The techniques described above for the classification of native valve condition were applied to the task of classifying the origin of the abnormal native valve population. Comparing the individual performance of the full 512 DWT coefficient features resulted in a maximum 84% correct classification being achieved. Using 37 top performers in the identification of optimum feature subsets it was found that by employing a four-feature subset all 25 population members could be correctly classified.
6.3.2 Feature subset selection and classification of aortic C-E bioprosthetic valve patient data

Prior to the identification of an optimum feature subset it is essential that initial dimensionality reduction is performed. Using the K-nearest neighbour classifier with 5 neighbours the performance of each of the 512 DWT coefficient features was analysed with respect to classification of the condition of aortic C-E bioprosthetic valve subjects. As in the native valve case a whole range of performances were achieved. Figure 6.12 shows a histogram of the number of features with respect to classification performance for the top 51 features.

![Histogram of Feature Performance](image)

Figure 6.12: Distribution of the top 51 classification performances for aortic C-E bioprosthetic valve subjects.

Inspecting Figure 6.12 and stipulating that a feature is only labelled as a top performer if it achieves a classification success of at least 16 (70% correct classification), it can be seen that all 51 features shown will be labelled as top performers. It can also be seen from Figure 6.12 that by using only individual features it is possible to correctly classify a maximum of 78% of the aortic C-E valve population. Figure 6.13 shows the position of these 51 top performing features on the time-scale plane. Also drawn on the time-scale plane of Figure 6.13 are the average positions of the R and T waves of the synchronous ECG signal.
Figure 6.13: Time-scale position of the 51 top performing features evaluated for the aortic 
C-E bioprosthetic valve population.

As in the native valve case, Figure 6.13 indicates that the C-E valve population top performing features are in part concentrated during systole. In addition to the systolic area of interest there does seem to be a number of features that occur after 700 ms, i.e. after the second heart sound early in the diastolic period. This area may be contributing to classification because of the occurrence of either a third heart sound, a fourth heart sound or an early diastolic murmur. Identification of any of these three sounds using auscultation is very difficult but it can be seen that using the optimised feature classification technique a pathological sound has been clearly identified. The prominent high-scale / low-frequency features from 800 - 900 ms in Figure 6.13 lend support to the possibility of a third heart sound influencing classification success.

The 51 top performing features were used to generate a covariance matrix from which 51 eigenvalues were extracted. The graph in Figure 6.14 shows the 51 eigenvalues ranked according to magnitude.
Chapter 6: Feature Extraction and Classification

Figure 6.14: Eigenvalues generated from the covariance matrix of the 51 top performing aortic C-E valve features.

The graph shows an obvious knee after eigenvalue 10. This suggests that the covariance matrix of the aortic C-E bioprosthetic valve population is rank deficient to a value of 10. Although this rank value is lower than that for the native valve case it is still high enough to suggest the possibility of an improved classification rate for an optimised feature subset.

An exhaustive search was performed on 2, 3 and 4 combinations of the feature set to identify the optimum feature subset that provided the highest classification rate. Due to the comparatively small size of the aortic C-E valve population it is not desirable to use feature subsets larger than 4. Table 6.6 shows the maximum level of correct classification achieved for each of the subset sizes.
Table 6.6: Optimised subset performance for classification of aortic C-E bioprosthetic valve condition.

The results in Table 6.6 show that a maximum of 87% correct classification was achieved for the identification of aortic C-E valve condition. It was found that the number of subsets achieving the maximum 87% classification rate increased from 2 to 11 between the 3 feature analysis and the 4 feature analysis. This apparent railing of the classification success rate suggests that 87% is a true rate for the separation of the C-E valve population when analysed using these techniques.

6.4 Summary

Table 6.7 shows the classification results presented in the previous sections of this chapter for all subject populations analysed using both feature extraction schemes.

Table 6.7: Compiled results of maximum classification rates achieved for the schemes and populations analysed in this chapter.
It can be seen from Table 6.7 that for every population tested, the optimum feature classification technique outperforms the equivalent morphological scheme. The inability of the morphological scheme to completely separate any of the test populations was not unexpected. After first generating the time-frequency distributions of all the test data a visual inspection of the various populations was performed in an attempt to identify if the data appeared separable and to identify features that looked to be promising class differentiators. The results of this visual inspection of the data suggested that separation of the data was not a trivial task. Hence, with the failure of the very efficient human pattern recognition system to classify the data it is unsurprising that an automated system using similar morphological features does not result in complete class separation. Using the blind DWT optimised feature extraction scheme allows the identification of a number of subtle features that have, as shown in Table 6.7, improved discriminative potential.

A particularly interesting feature in Table 6.7 is the contrast between the optimum and morphological results for the identification of dysfunctioning native valve origin. Comparing this contrast with the small rise in classification success in the results for identification of native valve condition it can be concluded that the morphological features listed in Table 6.1 are well suited to describing valve condition but not the origin of dysfunction. This result suggests that the identification of dysfunctioning valve origin requires an alternative feature set. Results in section 6.2.5 show that the two binary morphological features are useless in this classification task and hence, an improvement in results may be achieved with the identification of more suitable binary features.

A particular advantage of morphological features and a disadvantage of blind feature extraction is the possibility that classification is optimised to the available data rather than the true population. In studies where large amounts of test data is used it is valid to assume that results for the classification of the test set reflect the true population. In this study only small population sizes have been employed but there are a number of results that suggest that the true population has been represented. Firstly, as was demonstrated in section 5.3, there are a number of the C-E bioprosthetic valve population whose true class is debatable. It can be seen in Table 6.7 that using the optimum feature set to separate these classes only 87% were correctly classified. If the optimised scheme was wrongly performing a classification of the available data rather than the population one would expect a 100% classification rate. Further evidence of the validity of the optimised classification scheme was given in Figures 6.10 and 6.13 where it can be seen that the features used for classification have definite structure.
If classification were being performed on the data set rather than the true population it is quite possible that the top performing features would be randomly spread about the time-scale plane. The inherent structure of the optimised features and their strong relationship to the major sound components again supports the validity of the optimum classification technique.

The results in this section clearly show that non-stationary time-frequency and time-scale techniques can be used in the description and diagnosis of recorded heart sounds. The comparison of optimum and morphological schemes suggests that the classification of valve condition and origin is a non-trivial task. Results suggest that complete classification is only possible by extracting subtle features that may not be particularly prominent during a simple visual inspection. Of particular interest is the success of the optimised scheme in identifying an early diastolic pathological sound in the aortic C-E population. It is commonly documented [24] that such sounds are very difficult to identify during auscultation, hence, it can be concluded that the electronic recording and processing of sounds provides an attractive alternative to classical techniques. Also, the reduced optimum classification rate attained for the identification of C-E bioprosthetic valve condition together with the typical results shown in section 5.3 suggest that the optimised scheme has the potential to be used as an early indicator of valve dysfunction.
Chapter 7
Conclusions

This chapter provides a summary to the research presented in this thesis. In particular, comments are made as to the success of the work with relation to the aim and objectives first outlined in chapter 1. Details are given of the major achievements of the work and the contribution made during the period of study to knowledge in the emerging field of time-frequency analysis of heart sounds. Finally, a number of ideas are presented for improvement and further development of the research.

7.1 Summary and Conclusions of the Research

The first objective of the research was the collection of a significant data set comprising of subjects with various valve types in a number of alternative conditions and positions. The collection of a significant data set was felt to be an essential component of the research to allow a number of comparative results to be presented and definite conclusions to be made. In chapter 3 it was noted that under ideal situations the sample size used in an investigation should be a significant percentage of the population being analysed. It was also noted that the true limiting factors on the survey size analysed are availability of patients and the time available to conduct the survey. Under these constraints four major subject populations were recorded: native valve subjects, C-E bioprosthetic valve subjects, normal Bjork-Shiley metallic valve subjects and subjects before and after surgery for valve replacement. Considering this 100 patient test population it can be concluded that the first objective, to collect a significant data set, has clearly been achieved. In addition, from the detailed description of data acquisition equipment and procedures in chapter 3 it can also be concluded that the recorded heart sounds were of a high quality and provide a solid foundation for the research.
Prior to the transformation of any recorded data, a comparative analysis of four commonly used time-frequency techniques was performed. An initial indication of the performance of each transform was gained by applying each to two simple sinusoid test signals. Results generated from the analysis of a simple three sinusoid step signal indicated that the CWD produced a clearer and higher resolution description of the signal than any of the other transforms. In contrast, applying the four techniques to a simple dual chirp test signal resulted in the CWD performing particularly badly and the STFT producing the best signal description. From these contrasting results it can be concluded that the relative performance of the time-frequency transforms is almost entirely signal dependent. Previous research comparing the suitability of various time-frequency techniques to the analysis of heart sounds have presented results in terms of application to test signals and typical sound recordings. As already concluded, the analysis of constructed test signals only provides an academic indication of transform suitability to the task. Application of various transforms to a typical sound provides an indication of performance but without an a-priori knowledge of signal components it is impossible to quantify the integrity of the description. To overcome these limitations a simulated heart sound signal with known parameters was generated. From the analysis of transform results it was concluded that none of the four transforms provided an ideal description; the WT and STFT suffered from a lack of resolution, the WD suffered from extensive cross-term problems and the CWD exhibited small but detectable cross-terms. A comparison of results for the four transforms provided the conclusion that the CWD was the most suitable transform for the description and analysis of the recorded heart sounds. The rigorous analysis performed here brings into question the results presented by previous researchers that have concluded that the detailed dynamics of a heart sound can be investigated using RIDs such as the CWD. The results in this thesis suggest that further investigation of the possible appearance of undesirable cross-terms must be performed before such a conclusion can be made.

The second objective noted in chapter 1 was the comparative analysis of a number of alternative time-frequency techniques. From the discussion given above it can be seen that a clear conclusion as to the most suitable transform for the analysis of heart sounds has been achieved. It has also been noted that achieving this objective has raised an interesting question as to the suitability of RIDs such as the CWD to investigating the detailed dynamics of recorded heart sounds.

The third objective presented in chapter 1 was the analysis of the data set to investigate the effect valve type, condition and construction has on the appearance of the time-frequency
structure of the recorded heart sounds. By comparing typical and average results for contrast-
ing normal and abnormal valves it was concluded that there were a number of definite differ-
ences between normal and abnormal distributions. These results support the speculation of
previous researchers as to the suitability of time-frequency techniques to the description and
diagnosis of recorded heart sounds. In addition, it was found that abnormality of alternative
mitral and aortic valves was indicated in different ways and as such, results suggested a defi-
nite potential for not only the identification of valve integrity but also the origin of dysfunc-
tion. A particularly useful result demonstrated in all valve populations is the consistent
appearance of normal sounds. The results presented support the auscultation assumption that
the appearance of sounds other than the major first and second heart sounds is very often an
indication of dysfunction. Exceptions to this rule are particularly significant in the biopros-
thetic valve population, a number of the normal valve subjects have significant extra patho-
logical sounds. It is well documented that these bioprosthetic valves have relatively poor
durability and hence it is possible that these subjects are in an initial state of dysfunction.
When interviewed, all these subjects indicated that they were in a state of good health and
were not suffering from any obvious external symptoms such as shortness of breath. These
results would suggest that time-frequency techniques offer definite potential for the early
detection of valve dysfunction. As indicated in Figure 2.5, early detection of valve dysfunc-
tion offers the potential of significantly reducing undesirable patient mortality.

In addition to considering the effect valve condition has on the time-frequency structure of
recorded heart sounds, results also provide information as to the possible origins of the
sounds. From the analysis of the metallic valve subjects it was concluded that valve vibra-
tion occurs at both valve opening and closing but this valvular vibration plays only a small
role in the whole sound generation process. A comparison of recordings taken before and
after surgery for valve replacement indicate that valve condition has an effect on the whole
cardiac cycle. In previous research sound analysis has been performed on isolated first and
second sound segments. It can be concluded from the results in this work that to utilise all
available diagnostic information, the whole cardiac cycle should be considered. From the
results in chapter 5 a number of simple conclusions about the origins of sounds can be made.
Firstly, sounds are generated due to vibration of the whole cardiac system with a number of
structures and mechanisms providing a contribution. Secondly, the cardiac valves’ influence
upon the cycle of sounds is, in the majority, concentrated at onset of the first and second
sounds. Also, due to the system origin of sounds, a change in valve condition has an effect
on the whole system and hence, the whole cardiac cycle contains potential diagnostic infor-
mation.
Chapter 7: Conclusions

Considering the above discussion it can clearly be seen that the third objective presented in chapter 1 has been achieved. In achieving this objective it has been concluded that the time-frequency analysis of recorded heart sounds can be used as an indicator of valve integrity. In addition, considering the effect valve type and structure has upon the appearance of recorded sounds a hypothesis for the origin and nature of heart sounds has been developed.

Chapter 6 of the thesis describes the techniques used to perform an automated classification of the integrity of native and bioprosthetic valve populations. Two alternative methods for the extraction of features are presented: extraction of a set of morphological features and the extraction of a set of optimised features. Using the morphological feature set, classification rates of 96%, 72% and 83% were obtained for the identification of native valve condition, native valve origin of dysfunction and bioprosthetic valve condition respectively. In the case of the optimised feature classification, results for the three groups, as listed above, were 100%, 100% and 87% correct classification respectively. The classification task of identifying dysfunctioning valve origin refers to the identification, in known abnormal subjects, of whether the subject has a dysfunctioning aortic or mitral valve. It should be noted that in all past work, using stationary analysis techniques, subject diagnosis has been limited to identifying the condition of the cardiac valves irrespective of their position. Using non-stationary time-frequency analysis techniques it has been possible, in this work, to analyse the whole cardiac cycle of sounds. As a result it has been possible to extend the diagnosis past the identification of valve condition to the identification of valve condition and the origin of dysfunction. Considering the classification results it can be seen that using morphological features it is possible to perform a clear diagnosis of valve condition. In contrast, using the morphological features identified in chapter 6 it is only possible to provide an indication of valve origin for the dysfunctioning native population. From these contrasting results it can be concluded that the morphological feature set described in chapter 6 is particularly suited to the identification of valve condition but is ineffective as an indicator of the origin of valve dysfunction. Comparing classification results for morphological and optimised features it can be seen that in every population the optimised feature classification scheme outperforms the morphological classification scheme. This suggests that for a complete analysis of the data, a set of features that are possibly more subtle than those used in the morphological scheme are required.

From the discussions presented above it can be concluded that time-frequency analysis techniques can be used as a tool for the description and diagnosis of recorded heart sounds. It can also be concluded that using a non-stationary time-frequency technique rather than
traditional stationary frequency analysis procedures provides definite advantages due to the ability to consider all information in the whole cardiac cycle. Considering the whole cardiac cycle has avoided the need for a-priori knowledge of the position of dysfunction and hence, classification results are not only presented for valve condition but also the origin of dysfunction.

The overall aim of the work in this thesis has been the investigation of the suitability of modern time-frequency techniques to the analysis of surface recorded heart sounds. It can be concluded that using a high resolution time-frequency technique such as the CWD it is possible to generate a clear description of the structure of the cardiac sounds. The only limitation of such techniques is the possible appearance of undesirable transform cross-terms. These undesirable components bring into question the usefulness of time-frequency techniques for the detailed analysis of the dynamics of individual sounds. It can also be concluded that that the extraction of morphological features from the CWD of a recorded sound cycle can be used to identify valve integrity and provide an indication of the origin of valve dysfunction. Using the DWT to identify a set of optimum features it is possible to identify a set of non-stationary features that provide a clear indication of valve condition and origin.

7.2 Achievements

The first achievement detailed in this thesis is the collection of the data set. The populations detailed in chapter 3 are composed of 100 individual subjects each with a database of 8 signal recordings of PCG and ECG signals taken using different microphones in different positions on the subject's chest. Such a comprehensive database has only one or two equivalents in the world and hence the data represents a definite contribution to knowledge. It is envisaged that both the PCG and ECG data will be utilised extensively in future research. It should be noted that in total over 200 subjects were successfully recorded using the equipment and procedures described in chapter 3. These recordings were collected over a period of 2 years in an estimated total of 600 hours. Hence, with such a significant commitment of time and resources, it is unsurprising that the data set represents one of the largest of its kind in the world.

Chapter 4 includes a rigorous analysis and comparison of the performance of four common time-frequency techniques. As already discussed in the previous section, this analysis represents the first true attempt to quantitatively analyse the suitability of various time-frequency techniques to the analysis of heart sounds. Previous researchers have analysed typical
sounds and concluded that time-frequency techniques such as the CWD can be used for the
detailed description of individual sound dynamics. Results in this thesis suggest that further
investigation of the appearance of undesirable cross-terms should be performed before such a
conclusion can be made.

Since the commencement of this project a number of researchers have begun work on the
analysis of heart sounds using modern time-frequency techniques. These investigations have
concentrated on the analysis of the suitability of various advanced time-frequency techniques
to heart sound description. The work in this thesis represents the first publication of the use
of advanced time-frequency techniques for the analysis and diagnosis of a significant group
of cardiac patients.

As described in chapter 2 there is a definite lack of knowledge regarding the origins and gen-
erating mechanisms of heart sounds. In chapter 5 a number of observations were made about
the effect valve type, condition and structure have on the heart sound time-frequency distri-
bution. By considering these results a hypothesis was presented for the origin of sounds with
respect to the cardiac valves. The description presented is not the final solution to the prob-
lem but it does present a well supported foundation from which further investigation may be
performed.

As described in chapter 2 digital signal processing techniques have been applied to a great
number of heart sound analysis and diagnosis problems. In the past a number of researchers
have analysed the characteristics of normal native sounds but the comparison of normal and
abnormal native valves is an application that has to this date not been attempted. The results
in this thesis represent the first publication of the comparative analysis of normal and abnor-
mal native valve sounds using any spectral estimation technique. In addition to results being
presented for the analysis of native valve condition, the analysis is extended to include a clas-
sification of the origin of native valve dysfunction. To the author’s knowledge, the extension
of the diagnosis to the identification of origin of dysfunction has previously never been
attempted.

7.3 Further Work

From the results in section 4.6 it was concluded that all four transforms investigated did not
produce an ideal representation of the signal. In light of this it is proposed that the work in
Chapter 7: Conclusions

This thesis could be extended to the investigation and development of a time-frequency transform that produces a cross-term free, high resolution signal representation i.e. a zero interference distribution. Development of such a technique is far from trivial and presents a number of distinct theoretical and practical problems. A more practical aim may be the enhancement of present techniques to take into consideration signal characteristics. One possibility for such a signal optimised transform is the development of a mixed time-frequency transform. The joint application of two alternative transforms and the subsequent mixture of results is a concept that has already been applied in the spectral-domain. Hence, there may be potential for the implementation of similar techniques in the time-frequency domain. If the data is initially analysed using the STFT it should be possible to identify, due to the interference-free format of the STFT, areas in which signal components exist and areas that are free of such components. If the CWD is then used to transform the signal it should be possible to identify which terms in the transform are true signal components and which are cross-terms. In effect it is proposed that the STFT should be used to produce a low resolution, cross-term free time-frequency window function that can subsequently be applied to the CWD results to reduce undesirable cross-terms.

At the start of the project it was envisaged that the results of the research could be extended to the development of an "electronic stethoscope". If such an aim were to be achieved it would be necessary to increase the size of data set on which the techniques have been tested. There would have to be an extensive field survey to validate the usefulness and integrity of the equipment and techniques. In addition it would be necessary to analyse the performance of these techniques in identifying specific valvular problems such as stenosis, regurgitation and calcification. The results presented in this thesis have only provided an indication of the usefulness of time-frequency techniques and the complete validation of the techniques as a useful aid would require a substantial period of field study.

Results in chapter 6 indicate that using a blind optimised feature set it is possible to classify the data into their respective groups of valve condition and origin. The practical application of such a blind tool is limited, physicians invariably do not trust equipment that provides them with a diagnostic decision. The medical community often prefer electronic equipment that provides them with information from which a human decision can be made. In light of this it is envisaged that any further development of the work should be towards producing a visual aid that, due to the electronic acquisition of sound and the time-frequency visual presentation of information, provides an improvement on auscultation techniques. If this idea were taken further it is envisaged that further work would have to be performed on the
identification of morphological features that could be extracted visually to perform a diagnosis. The identification of these improved morphological features could be guided by the identification of optimum features using a blind classification scheme such as that described in section 6.3.
References


Bibliography


38. T. C. Hearn, J. Mazumdar, and R. Hubbard, "Temporal and Heart-Size Effects in First-

Fourier Transform in the Frequency Analysis of the Second Heart Sound in Normal

40. A. P. Yoganathan, R. Gupta, F. E. Udwadia, J. W. Miller, and W. H. Corcoran, "Use of
the Fast Fourier Transform for Frequency Analysis of the First Heart Sound in Normal

41. P. D. Stein, H. N. Sabbah, J. B. Lakier, and S. Goldstein, "Frequency Spectrum of the
Aortic Component of the Second Heart Sound in Patients With Normal Valves, Aortic
48-52 (1980).

42. P. D. Stein, H. N. Sabbah, J. B. Lakier, D. J. Magilligan, and S. Goldstein, "Frequency
of the First Heart Sound in the Assessment of Stiffening of Mitral Bioprosthetic

43. P. Stein, H. Sabbah, J. Lakier, S. Kemp, and D. Magilligan, "Frequency Spectra of First
Heart Sound and of Aortic Component of the Second Heart Sound in Patients with

44. Y. Kagawa, S. Nitta, N. Satoh, K. Saji, Y. Shibota, T. Horiuchi, and M. Tanaka, "Sound
Spectroanalytic Diagnosis of Malfunctioning Prosthetic Heart Valves," *Tohoku Journal

45. A. Iwata, N. Suzumura, and K. Ikegaya, "Pattern Classification of the Phonocardio-
gram using Linear Prediction Analysis," *Medical & Biological Engineering & Comput-


47. R. Bedi, "Signal Processing and Frequency Analysis of Carpentier-Edwards Biopros-

(1994).


Biomedical Engineering, 36 No 8, pp. 815-825 (1989).


Bibliography


Bibliography


Appendix A

List of Publications


†Reprinted in this appendix.
Abstract - This paper describes the novel application of time-frequency (T-F) distributions to the assessment of heart valve condition. The Choi-Williams distribution (CWD) was applied to a set of 45 normal/abnormal valve subjects and 9 descriptive features were extracted from the resulting T-F plots. A definition of normality, for each of the 9 features, was produced and assessing the performance of the 45 data subjects with respect to these criteria resulted in correct classification of 44 of the 45 subjects. This work empirically demonstrates the suitability of T-F techniques as an aid to the assessment of cardiac condition and clearly shows the advantages of T-F over classical spectral estimation techniques.

INTRODUCTION

Classical spectral estimation techniques have, in the past, been used as an indicator of cardiac valve condition [1]. More recently, research has moved towards investigating the suitability of advanced T-F techniques as descriptors of heart sound dynamics [2]. These works have all presented results as a comparison of the performance of various techniques when applied to simulated signals or a limited set of typical sound examples. This paper describes the application of the CWD to the description, feature extraction and ultimately, classification of a significant data set.

METHOD

Data Acquisition: The PCG and synchronous ECG were recorded from 45 subjects. This set included subjects with 10 normal mitral, 12 disfunctioning mitral, 10 normal aortic and 13 disfunctioning aortic valves. The PCG signal was recorded using a contact microphone (HP21050A) placed against the patient's chest, above the heart.

Choi-Williams Distribution: Cohen's general class of T-F distributions is defined in [3] as:

\[ P(t, f) = \int \int e^{j2\pi s(u-t)} \phi(v, \tau) s'(u-\tau/2) s(u+\tau/2)e^{-j2\pi f} dv df \]  

where \( s'(t) \) is the complex conjugate of \( s(t) \), the signal under analysis, and \( \phi(v, \tau) \) is an arbitrary function referred to as the kernel. Due to the bilinear nature of this general distribution, the analysis of multiple component signals results in the appearance of undesirable cross-terms. The reduction of these cross-terms by optimal choice of kernel function was first investigated by Choi and Williams [3]. They presented a new distribution based on the kernel function:

\[ \phi(\theta, \tau) = e^{\sigma^2r/\tau} \]  

where \( \sigma \) is a constant controlling the attenuation of cross-term components.

Feature Extraction and Classification: Using the CWD as a signal descriptor it is possible to simultaneously extract both time and frequency signal features. Table I describes the 9 features extracted from the T-F distribution generated, using the CWD, for each of the 45 subjects. S1 and S2 refer to the first and second sound respectively. The states of features 1 and 2 were determined via visual inspection and the more detailed features 3-9 were extracted automatically using a software based analysis tool.

Table I: Features extracted from subject T-F sound descriptions.

determined via visual inspection and the more detailed features 3-9 were extracted automatically using a software based analysis tool.

To assess the performance of the CWD as a suitable descriptor for heart sounds and the validity of the features listed in Table I, an attempt was made to separate the 45 data classes. The classification scheme used is...
RESULTS

Figure 1 shows the time-domain plot and CWD for a whole cardiac sound cycle recorded from a typical subject diagnosed as suffering from aortic valvular disease. The 9 features described in Table 1 were extracted from heart sounds. In comparison to past work, the results shown here highlight distinct advantages of T-F heart sound description over conventional spectral descriptions [1]. Firstly, the success of past work has relied on the extraction of complex signal features that have little physical interpretation. In this work classification was achieved with features that were a simple description of the T-F signal spectrum and hence, have a clear relation to the function of the cardiac system. Also, in past work using stationary analysis techniques, the cardiac cycle has been segmented into S1 and S2 pseudo-stationary periods and the remainder of the cardiac cycle discarded. In contrast, the methods described in this paper have been used to perform non-stationary signal analysis, which has allowed information available in the whole cardiac cycle to be used in the assessment of valve condition.

DISCUSSION AND CONCLUSIONS

Results show that the CWD is a valid tool for the description and pathological assessment of surface-recorded heart sounds. In comparison to past work, the results shown here highlight distinct advantages of T-F heart sound description over conventional spectral descriptions [1]. Firstly, the success of past work has relied on the extraction of complex signal features that have little physical interpretation. In this work classification was achieved with features that were a simple description of the T-F signal spectrum and hence, have a clear relation to the function of the cardiac system. Also, in past work using stationary analysis techniques, the cardiac cycle has been segmented into S1 and S2 pseudo-stationary periods and the remainder of the cardiac cycle discarded. In contrast, the methods described in this paper have been used to perform non-stationary signal analysis, which has allowed information available in the whole cardiac cycle to be used in the assessment of valve condition.
Wavelet transforms: an introduction

by P.M. Bentley and J.T.E. McDonnell

Wavelets and wavelet transforms are a relatively new topic in signal processing. Their development and, in particular, their application remains an active area of research. This paper presents a tutorial introduction to the theory, implementation and interpretation of the wavelet transform. The paper concentrates on the application of the wavelet transform to the time-scale (time-frequency) analysis of discrete signals. Examples are given of the analysis of basic test signals and of an actual electrocardiographic signal.

1 Introduction

Wavelet theory was initially proposed by the geophysicist J. Morlet and the theoretical physicist A. Grossmann. Together with their fellow Frenchman, Y. Meyer, this 'French school' developed the mathematical foundations of wavelets (ondelettes). At this stage wavelets were still very much in the realms of pure mathematics and, as such, concentrated more on the theory than the application. The two America-based researchers Daubechies and Mallat changed this by defining the connection between wavelets and digital signal processing.

Wavelets have been applied to a number of areas, including data compression, image processing and time-frequency spectral estimation. This paper, however, concentrates upon the application of the wavelet transform (WT) to the time-frequency analysis of discrete signals. As a time-frequency transform (the WT is more correctly described as a time-scale transform), the WT is a direct alternative to the short-time Fourier transform (STFT). By describing the theory and limitations of the STFT, the principles of time-frequency signal analysis are presented, after which the WT is described as a modification and alternative to the STFT. As mentioned above, the WT has its origins in the mathematical community and, as such, it is supported by extensive mathematical theory. In many ways the implementation of the WT is simpler than this theory would suggest. Hence, this paper includes only the minimum of mathematics necessary for understanding the WT and describes in detail the implementation and interpretation of the WT.

2 Time-frequency signal analysis and the STFT

In signal analysis few, if any, tools are as universal as the Fourier transform. It is used as the keystone of modern signal processing. The Fourier transform and its inverse are defined as follows:

\[
F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t) \, dt
\]

\[
f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(j\omega t) \, d\omega
\]

where \( F(\omega) \) is the Fourier transform of the signal \( f(t) \). Using the identity

\[
\exp(jk\theta) = \cos k\theta + j \sin k\theta
\]

the inverse transform can be described in terms of sine and cosine functions rather than complex exponentials:

\[
f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \cos (\omega t + j \sin \omega t) \, d\omega
\]

From this it can be seen that the Fourier transform \( F(\omega) \) of a signal is a function describing the contribution of sines and cosines to the construction of the original time domain signal. The contribution of these so-called basis functions is limited in time only by the duration of the signal being analysed.
The time independence of the basis functions of the Fourier transform results in a signal description purely in the frequency domain. The description of a signal as either a function of time or as a spectrum of frequency components contradicts our everyday experiences. The human auditory system relies upon both time and frequency parameters to identify and describe sounds. For example, to reproduce a piece of music, not only should all the correct notes be played but these notes should also be played in the correct order.

The Fourier transform is ideal for the analysis of stationary signals (signals whose statistical properties do not evolve with time). For the analysis of nonstationary signals a function is required that transforms a signal into a joint time–frequency domain. Such a description can be achieved using the well known STFT, which is an extension to the classical Fourier transform as defined by Gabor:

\[
STFT(t,w) = \int s(t) g(t - w) e^{-j2\pi ft} \, dt
\]

This function can be described as the Fourier transform of the signal \(s(t)\), previously windowed by the function \(g(t)\) around time \(t\). As the window function is shifted in time over the whole signal and consecutive overlapped transforms are performed, a description of the evolution of signal spectrum with time is achieved. This method assumes signal stationarity over the limited window \(g(t)\). If the window is relatively short this assumption of local stationarity is often valid. Eqn. 5 can be described diagrammatically as shown in Fig. 1. As the window function \(g(t)\) is shifted in time, repeated Fourier transforms are performed on overlapped data sets. Arranging these transforms chronologically and on a common frequency axis provides a time–frequency description of the signal which is commonly called the signal spectrogram.

Fig. 2c shows the result of applying the STFT, as described above, to a linear frequency modulated (LFM) test signal shown in Fig. 2a. Also shown in Fig. 2b is the Fourier transform of the LFM signal. Inspecting a signal as either a function of time or as its Fourier spectrum is inappropriate in many cases. Fig. 2 illustrates that for a complete description of the nonstationary LFM signal, its time–frequency description is far more appropriate.

The introduction of a window into the Fourier transform, as in the STFT, allows the generation of a time–frequency description of a signal. The introduction of such a window, however, has a detrimental effect upon frequency resolution. From Fig. 1 it can be seen that choosing a short window will result in a transform exhibiting good resolution in time. A short window also results in a reduced number of samples used in the Fourier transform calculation, which yields a reduced number of discrete frequencies that can be represented in the frequency domain. With a reduced number of discrete frequency intervals, the ability of the transform to discriminate between sinusoids of different frequencies is significantly

Fig. 3 Coverage of the time–frequency plane. The dimensions of the resolution squares represent minimum time and frequency intervals over which separate signals can be differentiated: (a) STFT; (b) wavelet transform.
reduced. The relationship between resolution in time and in frequency is a concept shared by many areas of science and is commonly referred to as the uncertainty principle. If \( \Delta t \) is the transform resolution in the time domain and \( \Delta f \) is the transform resolution in the frequency domain, then two sinusoids will be discriminated only if they are more than \( \Delta f \) apart in frequency or \( \Delta t \) apart in time, and the uncertainty principle can be written as

\[
\Delta t \Delta f \geq \frac{1}{4\pi}.
\]

Eqn. 6 implies that both time and frequency resolution cannot be made arbitrarily small: one must be traded for the other. The time and frequency resolution of the STFT is dependent upon the shape and length of the window function. Both these factors remain constant throughout an analysis and hence the time–frequency resolution also remains constant throughout an analysis. This constant joint time–frequency resolution results in the STFT covering the time–frequency plane with a uniform array of resolution squares, as shown in Fig. 3a. This trade-off between time and frequency resolution can be illustrated with a few simple examples.

Fig. 4 shows two simple test signals, both 256 ms long and sampled at a rate of 2000 samples per second. The test signal in Fig. 4a comprises two sinusoids of equal amplitude, one at 64 Hz and one at 192 Hz. The second test signal, in Fig. 4b, contains a single sinusoid with a 64-sample gap, during which the signal is 'switched off'. Fig. 5a shows the STFT of the test signal in Fig. 4a using a relatively long window length of 128 samples and Fig. 5b shows the STFT of the same signal with a shorter 32-sample window. The longer window length results in better frequency resolution and, as can be seen in Fig. 5a, the two sinusoids are clearly
resolved. The shorter window length results in reduced frequency resolution and this analysis, shown in Fig. 5b, fails to resolve the two sinusoidal components. Fig. 5 clearly demonstrates how transform window length affects resolution in frequency.

Fig. 6 shows the results of transforming the test signal in Fig. 4b, again using a long 128-sample window and a short 32-sample window. As would be expected, applying the longer 128-sample analysis (Fig. 6a) fails to resolve the signal gap but the short 32-sample analysis (Fig. 6b) quite clearly resolves the position and length of the gap. Fig. 5 and Fig. 6 together illustrate the joint time and frequency resolution limitations of the STFT implied by the uncertainty principle.

Although limited, the STFT is applicable to many problems where high-resolution characteristics are not required. When signal composition warrants high resolution or signal composition is unknown, alternative high-resolution techniques are required.

3 The wavelet transform

Previously, the STFT (eqn. 5) has been analysed as the Fourier transform of the windowed signal $s(t)g(t-r)$, but it is just as valid to describe this function as the decomposition of the signal $s(t)$ into the windowed basis functions $g(t-r)e^{in}$. The term ‘basis functions’ refers to a complete set of functions that can, when combined as a weighted sum, be used to construct a given signal. In the case of the STFT these basic functions are complex sinusoids, $e^{im}$, windowed by the function $g(t)$ centred around $r$. Using this description it is possible to write a general equation for the STFT in terms of basis functions $k_{nw}(t)$ and signal $s(t)$ as an inner product:

$$STFT(t,r) = \int s(t)k_{nw}(t)dt$$

For the STFT, the basis functions in eqn. 7 can be represented by $k_{nw}(t) = g(t)e^{in}$. Fig. 7a shows the real part
of three such functions to demonstrate the shape of typical STFT basis functions. These windowed basis functions are distinguished by their position in time $\tau$ and their frequency $\omega$. By mapping the signal onto these basis functions, a time-frequency description of the signal is generated.

The WT can also be described in terms of its basis functions, known as wavelets, using eqn. 7. In the case of the WT the frequency variable $\omega$ is replaced by the scale variable $a$ and generally the time-shift variable $\tau$ is represented by $b$. The wavelets are represented by

$$k_a(t) = \frac{1}{\sqrt{a}} h\left(\frac{t-b}{a}\right).$$

Substituting this description into eqn. 7 gives the definition for the continuous wavelet transform (CWT):

$$CWT(b,a) = \frac{1}{\sqrt{a}} \int h\left(\frac{t-b}{a}\right)s(t) \, dt$$

From this equation it can be seen that the WT performs a decomposition of the signal $s(t)$ into a weighted set of scaled wavelet functions $h(t)$. In general the wavelet $h(t)$ is a complex-valued function; Fig. 7b shows the real part of the Morlet wavelet at three different levels of scale. Comparing the two sets of basis functions in Fig. 7, it can be seen that the wavelets are all scaled versions of a common ‘mother wavelet’ whereas the basis functions of the STFT are windowed sinusoids.

Due to the scaling shown in Fig. 7b, wavelets at high frequencies are of limited duration and wavelets at low frequencies are relatively longer in duration. This varying window structure is reflected in the coverage of the time–frequency plane by the WT, as shown in Fig. 3b. These variable window length characteristics are obviously suited to the analysis of signals containing short high-frequency components and extended low-frequency components, which is often the case for signals encountered in practice. The WT employs a set of basis functions that are scaled versions of a single ‘mother function’. The ‘mother wavelet’ is constrained by a minimum of factors and, as a result, the function can take many forms. Fig. 8 shows a number of commonly used wavelets as cited in References 2, 3 and 5.

With such a varied array of possible wavelets, the choice of wavelet becomes an important factor. The best choice of wavelet for a particular application is unclear and is an active area of research.

Fig. 9 shows both the magnitude and phase plots generated by applying the WT, using the Morlet wavelet, to the LFM test signal in Fig. 2. Before describing the interpretation of Fig. 9, it is necessary to first describe the conventions used to graphically display the results of the WT.

The CWT as a time–scale transform has three dimensions. In this paper the three dimensions are represented on a $\log(a), b$ half-plane. The $\log(a)$ axis (scale) faces downwards and the $b$ axis (time-shift) faces to the right. The respective intensity of the transform at
The phase plot of a transformed signal does not show the same localisation as its equivalent modulus plot. Signal phase is a relative measure of complex and real components and, unlike modulus, it is not governed by localisation of signal intensity. For this reason the phase plot generated by the transformation of a real signal invariably covers the whole log(a),b plane. Such a plot is often difficult to interpret and, for this reason, the phase is selectively displayed. The phase is not displayed if the transform modulus drops below a predetermined cutoff value. Selective display of the phase plot is an aid to interpretation and for maximum clarity individual plots require individual thresholds.

The transformation of the LFM test signal in Fig. 9 clearly demonstrates these graphical conventions. Comparing the modulus plots of the LFM test signal for both the STFT and the wavelet transform, Fig. 2c and Fig. 9b respectively, it can be seen that both transforms produce similar results. The major difference between these two plots is that the wavelet transform results in a time-scale interpretation rather than the time-frequency description generated by the STFT. Inspecting the phase plot in Fig. 9c it can be seen that clear lines of constant phase are present. The separation of these constant phase lines is directly dependent upon signal frequency and independent of wavelet choice (see Section 5). By measuring the separation of these phase lines, signal frequency can be estimated and the relationship between the STFT time-frequency and the WT time-scale representations becomes apparent.

The STFT was originally introduced as a transform that performs signal analysis using a 'sliding' time window: the WT can also be described in this context. The distinct difference between each transform is that the STFT employs time windows of constant length over all frequency values and the CWT, due to the inclusion of the scale term a into the window function, employs time windows that decrease in length as frequency increases. By employing scaled window functions the WT does not overcome the uncertainty principle, but by employing variable window lengths, and hence variable resolution, an increase in performance may be achieved. Fig. 5 and Fig. 6 demonstrated the resolution limitations imposed upon the STFT by the uncertainty principle. Fig. 11 shows the magnitude and phase plots generated by applying the WT to the two test signals of Fig. 4. It can be seen from Fig. 11 that in both test cases the WT has clearly separated the signal components.

An important feature of Fig. 11b is the characteristic upturned V's that are present at the discontinuities in the interrupted sinusoid. It can be seen that the position of the signal discontinuities is clearly visible due to localisation in...
time at high frequency. This localisation of signal discontinuities is an important characteristic of the WT that can be used in signal interpretation.  

4 Implementation of wavelet transforms

The areas to which the WT is applied are varied and, as such, the WT has previously been implemented in a number of ways. This Section describes the two most common manifestations of the WT: the true discrete wavelet transform (DWT) used in image processing and data compression, and the oversampled WT used in time-scale spectral estimation.

Discretisation of the wavelet transform

Application of the WT in practice is achieved using digital computers to transform sampled signals. For this reason the CWT is replaced by the DWT. Implementation of the DWT can be introduced by considering subband decomposition — the digital filter equivalent of the DWT. The filter bank structure common to both subband decomposition and the DWT can be implemented efficiently using the tree structure shown in Fig. 12. Bandpass filtering is implemented as a lowpass—highpass filter pair which has mirrored characteristics. It is clear that it is impossible to use brickwall filters, but by using quadrature mirror filters (QMF) it is possible to closely approximate complete frequency coverage. At each stage of the tree structure in Fig. 12 the signal is split equally into its high- and low-frequency components. After filtering, the two components contain redundancies and it is valid to subsample each component by a factor of 2 without losing any information. Implementing this filter

---

**Fig. 11** (a) Magnitude and phase plot of the wavelet transform of a two-sinusoid test signal; (b) magnitude and phase plot of the wavelet transform of a sinusoid-with-gap test signal

**Fig. 12** Block diagram of the discrete wavelet transform implemented as subband decomposition using a filter bank tree
can be represented as a weighted sum of the basis functions and conversely that the signal can be perfectly reconstructed from a full set of weighted functions.

**Finer sampling in scale**

The octave band analysis, described above, has a number of applications, including data compression and image processing, but it does not generate results comparable to the spectrogram plots generated using the STFT. This is due to the efficient, but sparse, coverage of the time-frequency plane by the subsampled DWT. Fig. 14a shows the time-frequency sampling grid generated by the tree structure of Fig. 12. Subsampling by a factor of two at each octave level in Fig. 12 results in half the number of sample points in the consecutive octaves of Fig. 14a. To generate a scalogram plot comparable to the STFT spectrogram, a discretised version of the original CWT equation must be considered:

\[
\text{CWT}(f,r,a) = T_r \sum_{n=0}^{N} \left( \frac{n}{a} \right) f(n) \exp(-j2\pi nt)
\]

where \(n\) is the number of samples in the signal and \(T_r\) is the sampling interval. From this equation, it can be seen that at all values of the scale parameter \(a\), a full set of \(n\) samples is generated, i.e. no time domain subsampling takes place between octaves. It can also be seen that \(a\) is unconstrained, that is it can take any discrete value allowing the octave analysis to be subdivided. These octave subdivisions are often called voices. The values taken by the scale parameter \(a\) at discrete octave and voice levels of scale can be expressed as the series

\[
a = 2^{(m+M)j}
\]

where \(j\) is the octave number, \(m\) is the voice number and \(M\) is the number of voices per octave. Fig. 14b shows the time-scale grid generated using a number of voices per octave. Sample density along the horizontal time axis is governed by the sample frequency of the signal being transformed and down the vertical scale axis by the number of voices per octave. Using a suitably dense grid allows a scalogram with similar appearance to the typical STFT spectrogram to be produced.

**Computation load**

Although the direct implementation of the algorithm, as above, allows a clearer signal representation than the octave band DWT, this results in a higher computational load. The DWT implemented as the tree structure in Fig. 12 is efficient. It utilises common filter operations and performs dilation via simple signal decimation. In practice, \(L\) sample points analysed over \(J\) octaves require \(2L(1 - 2^{-J})\) multiplies/point and \(2L(1 - 1)(1 - 2^{-J})\) additions/point.\(^{1}\) In contrast, using the discretised CWT (eqn. 10) to produce a grid with \(M\) voices per octave, as in Fig. 14b, requires considerably more operations: \(2LJM\) multiplies/point and \(2MJ(L-1)\) additions/point.\(^{1}\) If a speech signal lasting a few seconds were to be transformed, a range of typically ten octaves along the scale axis and 8000 time-domain samples
per second would have to be analysed. To generate a clear plot as many as eight voices per octave might be used. Considering these typical values it can be seen that to generate transforms with a high degree of detail over extended periods would require an extensive amount of computation. This fact has been noted by a number of researchers and has lead to the development of a number of reduced computation algorithms[11,12] for calculating WTs.

One such algorithm is based upon the fast Fourier transform (FFT). By analysing the CWT in the Fourier domain the basic convolution operation of the CWT can be achieved via simple multiplication operations. Writing the CWT in the Fourier domain gives

\[ \text{FICWT}(b,a) = \mathcal{F}\left[ \mathcal{H}(\omega) \mathcal{S}(\omega) \right] \]  

where \( \text{FICWT}(b,a) \) is the Fourier transform of the continuous wavelet transform, \( \mathcal{H}(\omega) \) is the Fourier transform of the wavelet and \( \mathcal{S}(\omega) \) is the Fourier transform of the input signal. This equation can be used as the basis of an FFT-based fast wavelet transform. Eqn. 12 can be represented graphically as shown in Fig. 15.

Fig. 15 shows that the WT, after precalculation of the FFT of the wavelet and signal, is implemented via repeated scale, multiply and inverse FFT (IFFT) operations. This simple structure results in a complexity of \( \mathcal{O}(n \log n) \) multiplies/point and \( \mathcal{O}(n \log n) \) additions/point, where \( \log(2L) \) and split-radix implementation of the IFFT has been assumed.

Practically, the analysis of \( 2^n \) samples of speech (approximately 2 seconds at 8 kHz sampling rate) is over 700 times less complex when calculated using the fast wavelet transform algorithm rather than the direct implementation of the WT. Clearly the FFT-based fast wavelet transform and other equivalent fast algorithms provide significant computational savings. This not only allows the analysis of significant sample lengths but also real-time analysis becomes more realistic.

5 Selected examples

The WT is based upon very simple concepts but can produce results that, at first, appear to be difficult to interpret. The first example in this Section, the transformation of an impulse, is intended to illustrate how the WT can be used to interpret the time information in a signal. The second example, the transformation of a monochromatic signal, is intended to demonstrate how the WT represents the scale information in a signal and the relationship of this scale information to frequency. Finally, the transformation of a typical electrocardiogram (ECG) trace is presented to illustrate how the WT can be used to extract time, scale and frequency information from a practical signal.

Wavelet transform of an impulse

Considering the WT equation again:

\[ \text{CWT}(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h\left( \frac{t-b}{a} \right) s(t) \, dt \]  

To transform an impulse at time \( t_0 \), the above equation is evaluated at \( t_0 \) only:

\[ \text{CWT}(b,a) = \frac{1}{\sqrt{a}} h\left( \frac{t_0-b}{a} \right) \]  

Fig. 16 shows that the magnitude and phase of the Morlet wavelet

![Fig. 16 Magnitude and phase of the Morlet wavelet](image_url)
It can be seen that the resultant magnitude and phase terms of the transform are dependent upon magnitude and phase of the wavelet \( h^*(t) \). To predict the transform result, the magnitude and phase of the particular wavelet used must be known. In all the examples in this paper the Morlet wavelet is adopted:

\[
h(t) = \exp(j\omega_0 t) \exp(-\frac{t^2}{2})
\]  

(15)

Fig. 16 shows the magnitude and phase of this Morlet wavelet with \( \omega_0 \) set to 5.0.

Inspecting eqn. 14, it can be seen that at fixed values of the scale parameter \( a \), the transform magnitude is simply the magnitude of the wavelet centred about \( t \). As the scale parameter \( a \) increases, the magnitude of the transform still follows the magnitude of the wavelet but the wavelet will be scaled both in height and width. Fig. 17a shows the magnitude of the Morlet wavelet at a low scale value. Increasing the scale parameter \( a \) results in the magnitude of the wavelet being reduced in height and extended in length, as shown in Fig. 17b.

Fig. 18 shows the WT of an impulse using the Morlet wavelet. The high intensity of the magnitude plot at high frequencies and the spread occurring at low frequencies is concurrent with the change in wavelet magnitude for varying values of scale parameter \( a \). The phase plot generated by the WT of the impulse is again determined by the phase of the wavelet used. It can be seen from Fig. 16 that wavelet phase varies linearly with time but, as it rises to \( 2\pi \), wrap-around occurs and the phase drops to zero. This causes the linear increase in phase to be modulated by a sawtooth waveform, as in Fig. 18c. The results of Fig. 18 and the above discussion show the important influence that wavelet choice has upon results generated using the WT. In particular, it has been shown that both transform magnitude and phase are directly related to the wavelet used and as such analysis of transform results must take into consideration the wavelet function used. Fig. 18 also demonstrates the mechanisms behind the WT's ability to detect and localise signal discontinuities. Often these discontinuities carry the most important information in a signal. For example, the important information of the test signal of Fig. 4b may be the duration and position of the signal gap. Fig. 18b clearly demonstrates how the localisation about the signal discontinuities of the WT at low scale values (high frequency) can be used to highlight both edges of the signal gap.

Wavelet transform of a monochromatic signal

An example that illustrates clearly the relation between the WT as a time–scale analysis and the time–frequency interpretation of a signal is the WT of the real monochromatic signal \( s(t) = \cos(\omega_0 t) \). Writing the CWT in terms of Fourier transforms, where \( S(\omega) \) and \( H(\omega) \) are the Fourier transforms of the signal \( s(t) \) and the wavelet \( h(t) \), respectively, gives:

\[
CWT(b,a) = \omega_0 \int H'(\omega) e^{j\omega b} S(\omega) d\omega
\]

(16)

Substituting for \( s(t) \) gives a transform result that only exists at the two frequencies \( \omega_0 \) and \(-\omega_0 \). Therefore, the CWT is...
The two modulus terms in this function interfere and cause interpretation to become difficult. This problem will not arise if a progressive wavelet is used. A progressive wavelet is defined as a function that conforms to the basic wavelet conditions and exclusively comprises positive frequency components, i.e. for \( \omega < 0 \), \( \varphi(\omega) = 0 \). Using such a wavelet to transform the real monochromatic signal results in

\[
CWT(b, a) = \frac{1}{2} \exp(j\omega_0 b) H^*(\omega_0) + \frac{1}{2} \exp(-j\omega_0 b) H^*(-\omega_0)
\]  

(17)

The modulus of this transform, \( H^*(\omega_0) \), is independent of the variable \( b \), hence a plot of modulus will result in a uniform horizontal stripe on the log(\(a\)), \(b\) half-plane. This modulus term does have a relationship to the frequency of the signal being analysed but it is dependent upon the choice of wavelet. The phase term of the transform varies linearly with the variable \( b \) at a rate determined by the signal frequency. As the phase reaches \( 2\pi \), wrap-around occurs and the phase drops to zero. This causes the linear increase in phase to be modulated by a sawtooth waveform, which in turn results in distinct lines of constant phase. By measuring the separation of these lines of constant phase the frequency content of a signal can be easily interpreted. The result of using the progressive Morlet wavelet to transform a signal containing discrete frequency components has been illustrated previously in Fig. 11a. This example demonstrates how the modulus and phase information in a WT can be used to analyse the scale and hence frequency content of a signal.

Wavelet transform of a typical ECG

The final example of the application of the WT is the transformation of an electrocardiograph (ECG) trace. A typical ECG trace is shown in Fig. 19a. The magnitude and phase plots resulting from the transformation of this signal, again using the Morlet wavelet, are shown in Figs. 19b and c, respectively. The magnitude plot shows a large degree of spreading over the time-scale plane. This spreading may be explained in two ways. Localisation in the time-scale plane occurs when the signal and wavelet show a high degree of similarity. Comparing the ECG trace and the Morlet wavelet in Fig. 8, it can be seen that the two waveforms show little, if any, similarity. Hence, localisation of the transform is not expected. As seen from the first example, showing transformation of an impulse, the WT is particularly apt at highlighting discontinuities. Due to the sudden depolarisation and repolarisation of cardiac tissue the ECG contains a number of distinct peaks. It can be seen from the WT of the ECG that these peaks are highlighted as upturned V's, which become increasingly spread in time as scale increases. Although the impulsive nature of the ECG trace results in a spread transform at high scale values, it can be seen that at low values of scale, distinct time localisation occurs. This localisation characteristic can be used to pinpoint the significant peaks of the ECG trace. The position of significant complexes and peaks in the ECG waveform is an important diagnostic tool in detecting pathological conditions relating to the muscular and electrical tissue of the heart. These results point to the application of the WT to the automatic identification and characterisation of such pathological conditions.

6 Conclusions

By describing the relationship between the WT, the classical Fourier transform and its derivative the STFT, an
introduction to the theory of WT's has been presented. In the field of wavelets there are a great number of papers that describe the mathematical foundations of the subject. This paper is presented as an introduction to wavelets and as a primer to these more theoretical papers. Particular attention has been paid to the implementation of the WT in its true discrete form, the DWT, and by direct discretization: the CWT. Although the WT performs better in a number of situations than the STFT, this increase in performance is offset by a distinct increase in the complexity of the interpretation of the result. In an attempt to clarify the interpretation of results two simple examples have been presented: the WT of a monochromatic signal and the WT of an impulse. These two sets of results demonstrate the format in which both frequency and time information is represented by the WT. Finally, the transformation of a typical ECG trace demonstrates the ability of the WT to extract signal information from a nonstationary signal.

References

2. DAUBECHIES, I.: "Orthonormal bases of compactly supported wavelets", Communications on Pure and Applied Mathematics, 1988, 41, pp. 909-969
12. TCHAMITCHIAN, Ph.: "Wavelets, time-frequency methods and phase space" (Springer, Berlin, 1987, 2nd edn.), pp. 298-304

© IEE 1994
First received 19th January and in revised form 27th April 1994.

The authors are with the Department of Electrical Engineering, The University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JL, UK.

BOOK REVIEW

Mobile and personal communications
IEEE Conf., Pabl. 387
1993, 260pp., £47
ISBN: 0 85296 607 5

Developments in personal and mobile communications are so rapid that attendance at conferences is almost essential if one hopes to keep reasonably up-to-date. However, that is not always possible and next best is certainly to obtain a copy of the conference publication so that the papers of interest can be read.

Conferences, of course, cover a broad range of topics within the designated subject area and this one is no exception. Propagation, networks, system design, modulation, multiple access and private mobile radio are all featured as one might expect, and the authors are drawn from a wide variety of backgrounds, academic and industrial, from several different countries. Of particular interest in the proceedings of this conference is a section containing papers reporting activities within the European Communities' RACE programme. It seems that the investment in TDM as the multiple access system for networks of the future is now so large that any rival proposal (CDMA?) will have to show very substantial technical advances if it is to prevail.

All engineers involved in mobile and personal communication systems will find something of interest in this volume. There are 46 papers in total, providing the opportunity to read in depth those that are of specific interest whilst browsing through others to keep an awareness and an overall appreciation of where the subject is heading. At £47 it may not be a high priority for individual purchases, but it presents the state-of-the-art and it is well worth the effort of persuading your library to get a copy.

J.D. PARSONS